Exhibit 20

UNITED STATES PATENT AND TRADEMARK OFFICE

———————

BEFORE THE PATENT TRIAL AND APPEAL BOARD

———————

GOOGLE LLC,
Petitioner,

v.

SINGULAR COMPUTING LLC,
Patent Owner.

———————

IPR2021-00155
Patent 10,416,961 B2

———————

Before JUSTIN T. ARBES, STACEY G. WHITE, and JASON M. REPKO,
Administrative Patent Judges.

PER CURIAM.

JUDGMENT
Final Written Decision
Determining Some Challenged Claims Unpatentable
35 U.S.C. § 318(a)
Dismissing Patent Owner's Motion to Exclude
37 C.F.R. § 42.64
Granting Patent Owner's and Petitioner's Motions to Seal
37 C.F.R. §§ 42.14, 42.54

IPR2021-00155
Patent 10,416,961 B2

I.   INTRODUCTION

A.   Background and Summary

Petitioner Google LLC filed a Petition (Paper 2, "Pet.") requesting *inter partes* review of claims 1–5, 10, 13, 14, 21, and 23–25 of U.S. Patent No. 10,416,961 B2 (Ex. 1001, "the '961 patent") pursuant to 35 U.S.C. § 311(a). On May 12, 2021, we instituted an *inter partes* review as to all challenged claims on all grounds of unpatentability asserted in the Petition. Paper 16 ("Decision on Institution" or "Dec. on Inst."). Patent Owner Singular Computing LLC subsequently filed a Patent Owner Response (Paper 28, "PO Resp."), Petitioner filed a Reply (Paper 34, "Reply"), and Patent Owner filed a Sur-Reply (Paper 38, "Sur-Reply"). Patent Owner filed a Motion to Exclude (Paper 46, "Mot.") certain evidence submitted by Petitioner, to which Petitioner filed an Opposition (Paper 47, "Opp.") and Patent Owner filed a Reply (Paper 49, "Mot. Reply"). An oral hearing was held on February 11, 2022, and transcripts of the hearing are included in the record (Paper 55, "Conf. Tr."; Paper 56, "Public Tr."). The parties also filed Motions to Seal (Papers 52 and 54) portions of their demonstrative exhibits containing confidential information.

We have jurisdiction under 35 U.S.C. § 6. This Final Written Decision is issued pursuant to 35 U.S.C. § 318(a). For the reasons that follow, we determine that Petitioner has shown by a preponderance of the evidence that claims 1, 2, 4, 5, 10, 13, 14, 21, 24, and 25 are unpatentable, but has not shown by a preponderance of the evidence that claims 3 and 23 are unpatentable.

IPR2021-00155
Patent 10,416,961 B2

B. Related Matters

According to the parties, the '961 patent has been asserted in *Singular Computing LLC v. Google LLC*, No. 1:19-cv-12551-FDS (D. Mass.). *See* Pet. xi; Paper 6, 1 (Mandatory Notices).

Petitioner filed another petition challenging the '961 patent in IPR2021-00154, which was denied. *See* Pet. x; Paper 6, 1. The parties also identify the following *inter partes* reviews as related: IPR2021-00164 (denied) and IPR2021-00165 (instituted) challenging U.S. Patent No. 9,218,156 B2; and IPR2021-00178 (denied) and IPR2021-00179 (instituted) challenging U.S. Patent No. 8,407,273 B2. *See* Pet. xi; Paper 6, 1.

C. The '961 Patent

The '961 patent, entitled "Processing with Compact Arithmetic Processing Element," issued on September 17, 2019. Ex. 1001, code 45. The '961 patent relates to "computer processors or other devices which use low precision high dynamic range (LPHDR) processing elements to perform computations (such as arithmetic operations)." *Id.* at col. 5, ll. 58–62.
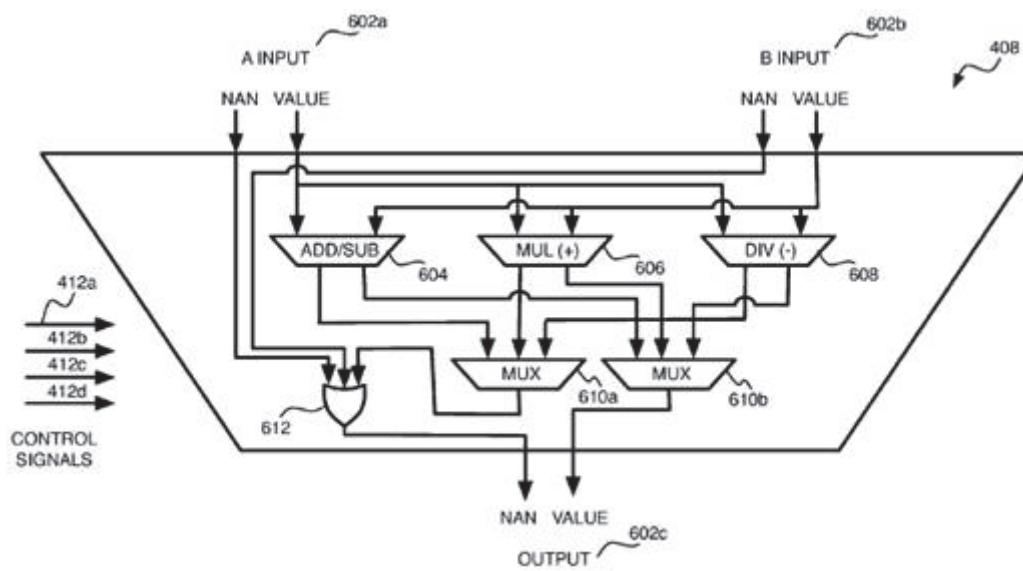
According to the '961 patent, conventional CPU chips make inefficient use of transistors as a tradeoff for delivering the high precision required by many applications. *Id.* at col. 2, l. 62–col. 3, l. 10. For example, conventional CPU chips "perform[] exact arithmetic with integers typically 32 or 64 bits long and perform[] rather accurate and widely standardized arithmetic with 32 and 64 bit floating point numbers," but require "on the order of a million transistors to implement the arithmetic operations." *Id.* at col. 3, ll. 3–10. According to the '961 patent, "many economically important applications . . . are not especially sensitive to precision and . . . would greatly benefit, in the form of application performance per transistor,

3

IPR2021-00155
Patent 10,416,961 B2

from the ability to draw upon a far greater fraction of the computing power inherent in those million transistors." *Id.* at col. 3, ll. 11–17. But "[c]urrent architectures for general purpose computing fail to deliver this power." *Id.* at col. 3, ll. 16–17.

The '961 patent is therefore "directed to a processor or other device, such as a programmable and/or massively parallel processor or other device, which includes processing elements designed to perform arithmetic operations . . . on numerical values of low precision but high dynamic range ('LPHDR arithmetic')." *Id.* at col. 1, l. 66–col. 2, l. 15. According to the '961 patent, "'low precision' processing elements perform arithmetic operations which produce results that frequently differ from exact results by at least 0.1%." *Id.* at col. 2, ll. 16–19. In addition, "high dynamic range" processing elements "are capable of operating on inputs and/or producing outputs spanning a range at least as large as from one millionth to one million." *Id.* at col. 2, ll. 23–27. Figure 6, reproduced below "is an example design for an LPHDR arithmetic unit according to one embodiment" of the '961 patent. Ex. 1001, 2:44–45.

4

IPR2021-00155
Patent 10,416,961 B2

FIG. 6 provides "an example design for an LPHDR arithmetic unit according to one embodiment of" the '961 patent. Ex. 1001, 2:44–45.

As shown in Figure 6, LPHDR arithmetic unit 408 receives two inputs: A input (602a) and B input (602b), and produces output 602c. *Id.* at col. 12, ll. 55–56. The LPHDR arithmetic unit "is controlled by control signals 412*a-d*, coming from the CU 106, that determine which available arithmetic operation will be performed on the inputs 602*a-b*." *Id.* at col. 12, ll. 62–65. According to the '961 patent, Figure 6 illustrates an embodiment where "all the available arithmetic operations are performed in parallel on the inputs 602*a-b* by adder/subtractor 604, multiplier 606, and divider 608." *Id.* at col. 12, l. 65–col. 13, l. 1. Finally, multiplexers (MUXes) 610a and 610b choose and send the desired result from among the outputs of the adder/subtractor, multiplier, and divider to output 602c. *Id.* at col. 13, ll. 5–14. The '961 patent provides that "[t]he computing architecture literature discusses many variations which may be incorporated into the embodiment illustrated in FIG. 6." *Id.* at col. 13, ll. 12–14.

According to the '961 patent, the "computational tasks" that the LPHDR arithmetic units can perform "enable a variety of practical applications." *Id.* at col. 17, ll. 30–33. The '961 patent provides, as examples, applications including "finding nearest neighbors," *id.* at col. 17, l. 41–col. 21, l. 43, "distance weighted scoring," *id.* at 21:45–22:35, and "removing motion blur in images," *id.* at col. 22, l. 37–col. 23, l. 48.

D. Claims

Of the claims challenged by Petitioner, claims 1, 10, and 21 are independent. Claim 1 is reproduced below.

5

IPR2021-00155
Patent 10,416,961 B2

1. A device comprising:

at least one first low precision high dynamic range (LPHDR) execution unit adapted to execute a first operation on a first input signal representing a first numerical value to produce a first output signal representing a second numerical value,

wherein the dynamic range of the possible valid inputs to the first operation is at least as wide as from 1/65,000 through 65,000 and for at least X=10% of the possible valid inputs to the first operation, the statistical mean, over repeated execution of the first operation on each specific input from the at least X % of the possible valid inputs to the first operation, of the numerical values represented by the first output signal of the LPHDR unit executing the first operation on that input differs by at least Y=0.2% from the result of an exact mathematical calculation of the first operation on the numerical values of that same input; and

at least one first computing device adapted to control the operation of the at least one first LPHDR execution unit.

Ex. 1001, col. 30, ll. 17–37.

E.   Evidence

The pending grounds of unpatentability in the instant *inter partes* review are based on the following prior art:

U.S. Patent No. 5,689,677, issued Nov. 18, 1997 (Ex. 1009, "MacMillan");

U.S. Patent Application Publication No. 2007/0203967 A1, published Aug. 30, 2007 (Ex. 1007, "Dockser"); and

Jonathan Ying Fai Tong, David Nagle, & Rob A. Rutenbar, "Reducing Power by Optimizing the Necessary Precision/Range of Floating-Point Arithmetic," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 8, no. 3 (June 2000) (Ex. 1008, "Tong").

Petitioner filed a declaration from Richard Goodin, P.E. (Ex. 1003, "Goodin Decl.") with its Petition and a reply declaration from Mr. Goodin (Ex. 1071)

IPR2021-00155
Patent 10,416,961 B2

with its Reply. Patent Owner filed declarations from Sunil P. Khatri, Ph.D. (Ex. 2051, "Khatri Decl.") and Joseph Bates, Ph.D. (Ex. 2052) with its Response. Also submitted as evidence are transcripts of the depositions of Mr. Goodin (Ex. 2043) and Dr. Khatri (Ex. 1072).

F. Asserted Grounds

This *inter partes* review involves the following grounds of unpatentability:

Claims Challenged	35 U.S.C. §	Reference(s)/Basis
1, 2, 4, 5, 10, 13, 14	103(a)[1]	Dockser
1, 2, 4, 5, 10, 13, 14, 21, 24, 25	103(a)	Dockser, Tong
1–5, 10, 13, 14	103(a)	Dockser, MacMillan
1–5, 10, 13, 14, 21, 23–25	103(a)	Dockser, Tong, MacMillan

II. ANALYSIS

A. Level of Ordinary Skill in the Art

In determining the level of ordinary skill in the art for a challenged patent, we look to "1) the types of problems encountered in the art; 2) the prior art solutions to those problems; 3) the rapidity with which innovations are made; 4) the sophistication of the technology; and 5) the educational level of active workers in the field." *Ruiz v. A.B. Chance Co.*, 234 F.3d 654,

---

[1] The Leahy-Smith America Invents Act, Pub. L. No. 112-29, 125 Stat. 284 (2011) ("AIA"), amended 35 U.S.C. § 103. Here, Petitioner's challenges are based on the pre-AIA version of 35 U.S.C. § 103.

7

IPR2021-00155
Patent 10,416,961 B2

666–67 (Fed. Cir. 2000). "Not all such factors may be present in every case, and one or more of them may predominate." *Id.*

Petitioner asserts that at the time of the earliest possible effective filing date of the '961 patent (June 19, 2009), a person of ordinary skill in the art would have had "at least a bachelor's degree in Electrical Engineering, Computer Engineering, Applied Mathematics, or the equivalent, and at least two years of academic or industry experience in computer architecture." Pet. 9 (citing Goodin Decl. ¶ 43). Patent Owner's proposed definition of the level of ordinary skill in the art is the same other than it removes the phrases "or the equivalent" and "at least." PO Resp. 14 (citing Khatri Decl. ¶¶ 36–37). There is little difference between the parties' proposed definitions, and the parties and their experts do not explain the basis for their proposals. Arguably, however, the term "at least" creates unnecessary ambiguity. Based on the full record developed during trial, including our review of the '961 patent and the types of problems and solutions described in the '961 patent and cited prior art, we determine that a person of ordinary skill in the art would have had a bachelor's degree in Electrical Engineering, Computer Engineering, Applied Mathematics, or the equivalent, and two years of academic or industry experience in computer architecture. *See, e.g.*, Ex. 1001, col. 1, l. 15–62 (describing in the "Background" section of the '961 patent various conventional methods of computation and their alleged deficiencies). We apply that definition for purposes of this Decision.

IPR2021-00155
Patent 10,416,961 B2

B. Claim Interpretation

We need only construe terms that are in controversy and only to the extent necessary to resolve the controversy. *Nidec Motor Corp. v. Zhongshan Broad Ocean Motor Co.*, 868 F.3d 1013, 1017 (Fed. Cir. 2017).

Claim 1 recites, in part, "at least one first low precision high dynamic range (LPHDR) execution unit adapted to execute a first operation on a first input signal representing a first numerical value to produce a first output signal representing a second numerical value." Ex. 1001, col. 30, ll. 18–22. Independent claims 10 and 21 recite similar limitations. *See id.* at col. 31, ll. 27–31 (claim 10); col. 32, ll. 43–47 (claim 21).

At issue here is the meaning of "low precision high dynamic range (LPHDR) execution unit," as recited in claims 1, 10, and 21. Specifically, we need to resolve the question of whether this execution unit must be adapted to execute arithmetic operations *only* at low precision. For the reasons that follow, we determine that it does not.

Petitioner argues that, in the prior art used in all challenges, Dockser's floating-point processor (FPP), including the floating-point operator (FPO) inside the FPP, is low precision "because 'the precision' of operations in the FPP is 'reduced' . . . , and because it operates with the minimum imprecision" recited in the claim. Pet. 15 (citing Goodin Decl. ¶¶ 208–209; Ex. 1001, col. 26, l. 61–col. 28, l. 8; Ex. 1035, 8). Petitioner asserts that, in the claim, the minimum imprecision is found in the limitation that it calls *1B2*:

> for at least X=10% of the possible valid inputs to the first operation, the statistical mean, over repeated execution of the first operation on each specific input from the at least X % of the possible valid inputs to the first operation, of the numerical values represented by the first output signal of the LPHDR unit

9

IPR2021-00155
Patent 10,416,961 B2

> executing the first operation on that input differs by at least
> Y=0.2% from the result of an exact mathematical calculation of
> the first operation on the numerical values of that same input.

Ex. 1001, col. 31, ll. 34–43. According to Petitioner, the '961 patent
characterizes degrees of precision described in language mirroring *1B2* as
"low precision." Pet. 15 (citing Ex. 1001, col. 26, l. 61–col. 28, l. 8).
Petitioner argues that the execution units in the prior art fall within the scope
of the recited unit if they perform reduced precision operations and at least
one, but not necessarily all, of those operations meets the other requirements
of claim 1. *See, e.g.*, *id.* at 32 (explaining that "numerous obvious
implementations of Dockser's FPP drop sufficient bits to yield an execution
unit meeting the claimed minimum imprecision"); Reply 2–3 (discussing the
"low precision" requirement).

Patent Owner argues that the LPHDR execution unit should be
construed as "an execution unit that executes arithmetic operations *only* at
low precision and with high dynamic range, wherein 'high dynamic range'
and 'low precision' are defined according to the numerical requirements
below." PO Resp. 15 (emphasis added). We emphasize "only" because
Patent Owner argues the recited LPHDR execution unit "necessarily
excludes full-precision or mixed full and low precision units." *Id.* at 16
(citing Khatri Decl. ¶ 60).

In support of its proposed construction, Patent Owner asserts that
construing the LPHDR execution unit as encompassing full or mixed
precision units would "impermissibly read out 'low precision high dynamic
range' and render the limitation meaningless." *Id.* (citing *Network-1 Techs.,
Inc. v. Hewlett-Packard Co.*, 981 F.3d 1015, 1022–23 (Fed. Cir. 2020)); *see*

IPR2021-00155
Patent 10,416,961 B2

also id. at 17 (arguing Petitioner's construction reads "'low precision' out of the claims entirely").

Based on the totality of the record, we agree with Petitioner that the recited LPHDR execution unit does not exclude mixed full and low precision execution units (i.e., those that are capable of arithmetic operations at both levels of precision). *See, e.g.*, Pet. 15, 32. We reject Patent Owner's argument that interpreting the claim in this way renders the term "low precision" "meaningless." PO Resp. 15–16. Rather, the term "low precision" imposes the requirement that the unit must *at least* perform arithmetic operations at low precision, which would exclude units that operate only at full precision, for example. Thus, we disagree with Patent Owner and assign little weight to the Khatri Declaration on this issue. *See id.* at 15–17; Khatri Decl. ¶¶ 60–61.

As stated in our Decision to Institute, the recitation of "a first operation" in claim 1 requires only one or more first operations that meet the low precision criteria specified in the claim. Dec. on Inst. 18. The claims do not require that "every" operation must be low precision or exclude the LPHDR execution unit from having other capabilities for other operations. *Id.* Indeed, Patent Owner agrees that "a claim's recitation of an 'execution unit adapted to execute a first operation' would normally not preclude that execution unit from performing other types of operations." PO Resp. 17. Patent Owner, though, argues that this "rule does not apply when the specification teaches otherwise, as is the situation here." *Id.* at 17–18. We disagree.

Patent Owner relies on exemplary non-limiting embodiments from the '961 patent. *See id.* at 18. Specifically, Patent Owner argues that "the claimed LPHDR execution units are smaller and have fewer transistors" than

11

IPR2021-00155
Patent 10,416,961 B2

prior art full precision execution units. *Id.* (citing Khatri Decl. ¶ 62). But

the cited passage pertains only to exemplary embodiments and says that the

"amount of resources" is small, with transistors being only one example of

such resources:

> For example, embodiments of the present invention may
> be implemented as any kind of machine which uses LPHDR
> arithmetic processing elements to provide computing using a
> *small amount of resources (e.g., transistors or volume)* compared
> with traditional architectures.

Ex. 1001, 8:1–5 (emphasis added). In fact, some embodiments do not even

require transistors: "Examples of such technologies include . . . other

technologies whether based on transistors or not . . . ." Ex. 1001, col. 26, ll.

27–42, *cited in* Reply 6.

The reference to the "small amount of resources" in the cited passage

(Ex. 1001, col. 8, ll. 1–5) is consistent with Petitioner's construction that the

LPHDR execution unit need only perform low precision operations some of

the time because the record shows that doing so would save power and other

computational resources. For example, Dr. Khatri testified that "exact

computing . . . takes up a lot of resources, takes up a lot of power, takes up a

lot of area, and takes up a lot of delay." Ex. 1072 ("Deposition of Sunil P.

Khatri"), 47:2–5. Mr. Goodin explained that reducing precision may

achieve power savings. *See, e.g.*, Ex. 2043 ("Deposition of Richard M.

Goodin"), 34:21–35:4; *see also id.* at 31:2–32:1 (discussing how selectable

precision may achieve savings). The Goodin Declaration cites specific

examples of how this is achieved. *See, e.g.*, Goodin Decl. ¶¶ 31–33 (citing

Ex. 1008, 273, 277–279); ¶ 35 (citing Ex. 1007 ¶¶ 3–7). Thus, we disagree

with Patent Owner's argument about reducing transistors and assign little

12

IPR2021-00155
Patent 10,416,961 B2

weight to the Khatri Declaration on this issue. *See* PO Resp. 18; Khatri Decl. ¶ 62.

Also, we agree with Petitioner that the degree of low precision is described in limitation *1B2* in claim 1 and similar limitations in the other claims. Pet. 15. The language used in *1B2* is consistent with how the patent describes low precision. Ex. 1001, col. 26, l. 61–col. 28, l. 8, *cited in* Pet. 15. For example, the patent states, "[t]he degree of precision of a 'low precision, high dynamic range' arithmetic element may vary from implementation to implementation." *Id.* at col. 26, ll. 61–63. The patent then lists several examples. *Id.* at col. 26, l. 61–col. 28, l. 8. In one example, "a LPHDR arithmetic element may produce results which are sometimes (or all of the time) no closer than 0.2% to the correct result." *Id.* at col. 27, ll. 7–10. Similarly, the claims recite, "the numerical values represented by the first output signal of the LPHDR unit executing the first operation on that input differs by at least $Y=0.2\%$ from the result of an exact mathematical calculation of the first operation on the numerical values of that same input." *Id.* at col. 30, ll. 29–34 (claim 1); col. 31, ll. 38–43 (claim 10); col. 32, ll. 54–59 (claim 21).

Patent Owner agrees that the claims provide the parameters of the low precision and high dynamic range in the limitations "wherein the dynamic range . . ." and "for at least $X=10\%$ . . . ." PO Resp. 16. Indeed, Patent Owner in the related district court case argued that the subsequent claim language specifying percentages "*defines* the term 'low precision'" and "specifies the degree of precision that an LPHDR execution unit must have in order to satisfy the 'low precision' . . . limitation of the claim." Ex. 1065, 9 (emphasis added); *see also* Ex. 1066, 12 (arguing that "'low precision high dynamic range' is defined in [the] claim itself"); Reply 2. Patent Owner,

IPR2021-00155
Patent 10,416,961 B2

however, argues that those limitations "do not allow for the execution unit to have full precision (or [low] dynamic range) capability." PO Resp. 16.

But the claims do not recite that every operation must meet these criteria. Rather, the claims only recite that the criteria apply to "a first operation." Ex. 1001, col. 30, ll. 23–34. This reading is consistent with the examples of LPHDR elements from the written description. *See* Ex. 1001, col. 26, l. 61–col. 28, l. 8, *cited in* Pet. 15. In those examples, the LPHDR arithmetic elements may produce results that are only "sometimes" no closer than the percentage to the correct result. *See id.* Thus, we agree with Patent Owner's argument that the claims provide the parameters for the term "low precision," but we disagree that the parameters must apply to every operation. *See* PO Resp. 16. For the same reasons, we assign little weight to the Khatri Declaration on this issue. *See* Khatri Decl. ¶¶ 60–64.

Patent Owner also argues that "small size and low transistor count . . . is what allows a much larger number of LPHDR execution units to operate in parallel on a single chip." PO Resp. 19. But the requirement of a "much larger number" is not in the independent claims. Nor do the claims (or Patent Owner's proposed construction) include any limitations on size of the LPHDR execution unit or its transistor count. And "not every benefit flowing from an invention is a claim limitation." *i4i Ltd. P'ship v. Microsoft Corp.*, 598 F.3d 831, 843 (Fed. Cir. 2010). Even so, the specification states that some embodiments are not based on transistors. *See, e.g.*, Ex. 1001, col. 26, ll. 27–42. So we find no support in the written description for Patent Owner's argument about small size and low transistor count. *See* PO Resp. 19. For the same reasons, we assign little weight to the Khatri Declaration on this issue. *See* Khatri Decl. ¶¶ 63–64.

IPR2021-00155
Patent 10,416,961 B2

We also agree with Petitioner that the written description does not support Patent Owner's assertion that the patent contrasts LPHDR execution units with other units that operate in both high and low precision. PO Resp. 19, 23 (citing Ex. 1001, 5:23–33); Reply 7 (discussing Patent Owner's argument). The relevant part of the patent describes a graphics processor that supports 16-bit and 32-bit floating point formats:

> When a graphics processor includes support for 16 bit floating point, that support is alongside support for 32 bit floating point, and increasingly, 64 bit floating point. That is, the 16 bit floating point format is supported for those applications that want it, but the higher precision formats also are supported because they are believed to be needed for traditional graphics applications and also for so called "general purpose" GPU applications. Thus, existing GPUs devote substantial resources to 32 (and increasingly 64) bit arithmetic and are wasteful of transistors in the sense discussed above.

Ex. 1001, col. 5, ll. 23–33. This paragraph does not define the term "low precision." *Id.* Nor does the patent explain how the 16-bit units meet the claimed requirements for low precision units that are recited in the claim. *See id.* Also, there is no discussion of these units in the part of the written description that lists examples of low precision units. *See id.* at col. 26, ll. 61–col. 28, ll. 8. Thus, we disagree with Patent Owner that the patent contrasts LHPDR units with those that operate in both high and low precision. PO Resp. 19, 23. For the same reasons, we assign little weight to the Khatri Declaration on this issue. *See* Khatri Decl. ¶¶ 63–64.

To the contrary, we agree with Petitioner that the patent discloses at least one embodiment that performs operations other than LPHDR computations:

> [W]hile the obvious method of using the above [logarithmic number system (LNS)] operations is to do LPHDR

15

IPR2021-00155
Patent 10,416,961 B2

> arithmetic, the programmer also may consider selected values to
> be 12 bit two's complement binary numbers. MUL and DIV may
> be used to add and subtract such values . . . . *So besides doing*
> *LPHDR computations*, this digital embodiment using LNS can
> perform simple binary arithmetic on short signed integers.

Ex. 1001, col. 14, ll. 12–22 (emphasis added), *quoted in* Reply 4. Because

the unit performs arithmetic operations other than those at low precision, this

example contradicts Patent Owner's argument that an LPHDR execution

unit is "an execution unit that executes arithmetic operations *only* at low

precision." PO Resp. 15 (emphasis added).

Patent Owner counters, "as Dr. Khatri explained, one of skill in the art

would not apply the terms 'precision' or 'dynamic range' to integers." Sur-

Reply 4 (citing Ex. 1072, 116:5–118:3). In Patent Owner's view, a person

of ordinary skill in the art "would understand that [Patent Owner]'s proposed

construction would *not* exclude execution units that can perform integer

operations, like the LNS embodiment described in the patent specification."

Id. at 4–5. Yet Patent Owner's proposed construction makes no distinction

between integer or floating-point operations. PO Resp. 15. Rather, the

construction refers to "arithmetic operations," generally. *See id.*

In its Sur-Reply, Patent Owner proposes an alternative construction—

for the first time in this proceeding—that further defines "arithmetic

operations" to be "*floating-point* arithmetic operations." Sur-Reply 5

(emphasis added). But a sur-reply must only respond to arguments raised in

the preceding brief. Patent Trial and Appeal Board, *Consolidated Trial*

Practice Guide 73–74 (Nov. 2019) ("TPG").[2] Under 37 C.F.R. § 42.23(b),

"respond" does not mean proceed in a new direction with a new approach.

---

[2] https://www.uspto.gov/sites/default/files/documents/tpgnov.pdf

IPR2021-00155
Patent 10,416,961 B2

See id. at 74. "[A] sur-reply that raises a new issue or belatedly presents evidence may not be considered." *Id.* So, apart from noting that we do not see any reason to limit the LPHDR execution unit to only low precision operations—regardless of how those operations are defined—for the reasons discussed above, we do not further consider this new claim construction presented for the first time in the Sur-Reply. *See* Sur-Reply 5.

In sum, we agree with Petitioner that the recited "low precision high dynamic range (LPHDR) execution unit" does not exclude units that are capable of performing operations at low and full precision, and the degree of "low precision" is defined in limitation *1B2*. *See, e.g.*, Pet. 15, 32; Reply 2–3. To determine the patentability of the challenged claims, we need not further construe LPHDR execution unit. *See Nidec*, 868 F.3d at 1017.

C. Legal Standards

To prevail in its challenges to the patentability of claims 1–5, 10, 13, 14, 21, and 23–25 of the '961 patent, Petitioner must demonstrate by a preponderance of the evidence that the claims are unpatentable. 35 U.S.C. § 316(e). "In an [*inter partes* review], the petitioner has the burden from the onset to show with particularity why the patent it challenges is unpatentable." *Harmonic Inc. v. Avid Tech., Inc.*, 815 F.3d 1356, 1363 (Fed. Cir. 2016). This burden of persuasion never shifts to Patent Owner. *Dynamic Drinkware, LLC v. Nat'l Graphics, Inc.*, 800 F.3d 1375, 1378 (Fed. Cir. 2015); *see also In re Magnum Oil Tools Int'l, Ltd.*, 829 F.3d 1364, 1376 (Fed. Cir. 2016) ("Where, as here, the only question presented is whether due consideration of the four *Graham* factors renders a claim or claims obvious, no burden shifts from the patent challenger to the patentee.").

IPR2021-00155
Patent 10,416,961 B2

A claim is unpatentable for obviousness if, to one of ordinary skill in the pertinent art, "the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made." *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 406 (2007) (quoting 35 U.S.C. § 103(a) (2006)). The question of obviousness is resolved on the basis of underlying factual determinations, including "the scope and content of the prior art"; "differences between the prior art and the claims at issue"; and "the level of ordinary skill in the pertinent art." *Graham v. John Deere Co.*, 383 U.S. 1, 17–18 (1966). Additionally, objective indicia of nonobviousness, such as "commercial success, long felt but unsolved needs, failure of others, etc., might be utilized to give light to the circumstances surrounding the origin of the subject matter sought to be patented. As indicia of obviousness or nonobviousness, these inquiries may have relevancy." *Id.* When conducting an obviousness analysis, we consider a prior art reference "not only for what it expressly teaches, but also for what it fairly suggests." *Bradium Techs. LLC v. Iancu*, 923 F.3d 1032, 1049 (Fed. Cir. 2019) (citation omitted).

A patent claim "is not proved obvious merely by demonstrating that each of its elements was, independently, known in the prior art." *KSR*, 550 U.S. at 418. An obviousness determination requires finding "both 'that a skilled artisan would have been motivated to combine the teachings of the prior art references to achieve the claimed invention, and that the skilled artisan would have had a reasonable expectation of success in doing so.'" *Intelligent Bio-Sys., Inc. v. Illumina Cambridge Ltd.*, 821 F.3d 1359, 1367–68 (Fed. Cir. 2016) (citation omitted); *see KSR*, 550 U.S. at 418 (for an obviousness analysis, "it can be important to identify a reason that would have prompted a person of ordinary skill in the relevant field to

IPR2021-00155
Patent 10,416,961 B2

combine the elements in the way the claimed new invention does"). Also,

"[t]hough less common, in appropriate circumstances, a patent can be

obvious in light of a single prior art reference if it would have been obvious

to modify that reference to arrive at the patented invention." *Arendi S.A.R.L.*

v. Apple Inc., 832 F.3d 1355, 1361 (Fed. Cir. 2016).

 "Although the *KSR* test is flexible, the Board 'must still be careful not

to allow hindsight reconstruction of references . . . without any explanation

as to *how* or *why* the references would be combined to produce the claimed

invention.'" *TriVascular, Inc. v. Samuels*, 812 F.3d 1056, 1066 (Fed. Cir.

2016) (citation omitted). Further, an assertion of obviousness "cannot be

sustained by mere conclusory statements; instead, there must be some

articulated reasoning with some rational underpinning to support the legal

conclusion of obviousness." *KSR*, 550 U.S. at 418 (quoting *In re Kahn*,

441 F.3d 977, 988 (Fed. Cir. 2006)); *accord In re NuVasive, Inc.*, 842 F.3d

1376, 1383 (Fed. Cir. 2016) (stating that "conclusory statements" amount to

an "insufficient articulation[] of motivation to combine"; "instead, the

finding must be supported by a 'reasoned explanation'" (citation omitted));

Magnum Oil, 829 F.3d at 1380 ("To satisfy its burden of proving

obviousness, a petitioner cannot employ mere conclusory statements. The

petitioner must instead articulate specific reasoning, based on evidence of

record, to support the legal conclusion of obviousness.").

 D. Obviousness Ground Based on Dockser (Claims 1, 2, 4, 5, 10, 13, and 14)

 Petitioner contends that claims 1, 2, 4, 5, 10, 13, and 14 are

unpatentable over Dockser under 35 U.S.C. § 103(a). Pet. 45–56.

IPR2021-00155
Patent 10,416,961 B2

1.   Dockser

Dockser discloses performing floating-point operations with a floating-point processor having "selectable subprecision." Ex. 1007, code (57), ¶¶ 15, 17. "A floating-point representation of a number commonly includes a sign component, an exponent, and a mantissa. To find the value of a floating-point number, the mantissa is multiplied by a base (commonly 2 in computers) raised to the power of the exponent. The sign is applied to the resultant value." *Id.* ¶ 1. "The precision of the floating-point processor is defined by the number of bits used to represent the mantissa. The more bits in the mantissa, the greater the precision." *Id.* ¶ 2. "The precision of [a] floating-point processor generally depends on the particular application. For example, the ANSI/IEEE-754 standard (commonly followed by modern computers) specifies a 32-bit single format having a 1-bit sign, an 8-bit exponent, and a 23-bit mantissa." *Id.* "Higher precision results in a higher accuracy, but commonly results in increased power consumption." *Id.* Dockser explains that the performance of floating-point operations also can be computationally inefficient because "[w]hile some applications may require [high] types of precision, other applications may not." *Id.* ¶ 3. For example, "some graphics applications may only require a 16-bit mantissa," such that "any accuracy beyond 16 bits of precision tends to result in unnecessary power consumption," but other applications may require "greater precision." *Id.* Accordingly, there was "a need in the art for a floating-point processor in which the reduced precision, or subprecision, of the floating-point format is selectable." *Id.*

Figure 1 of Dockser is reproduced below.
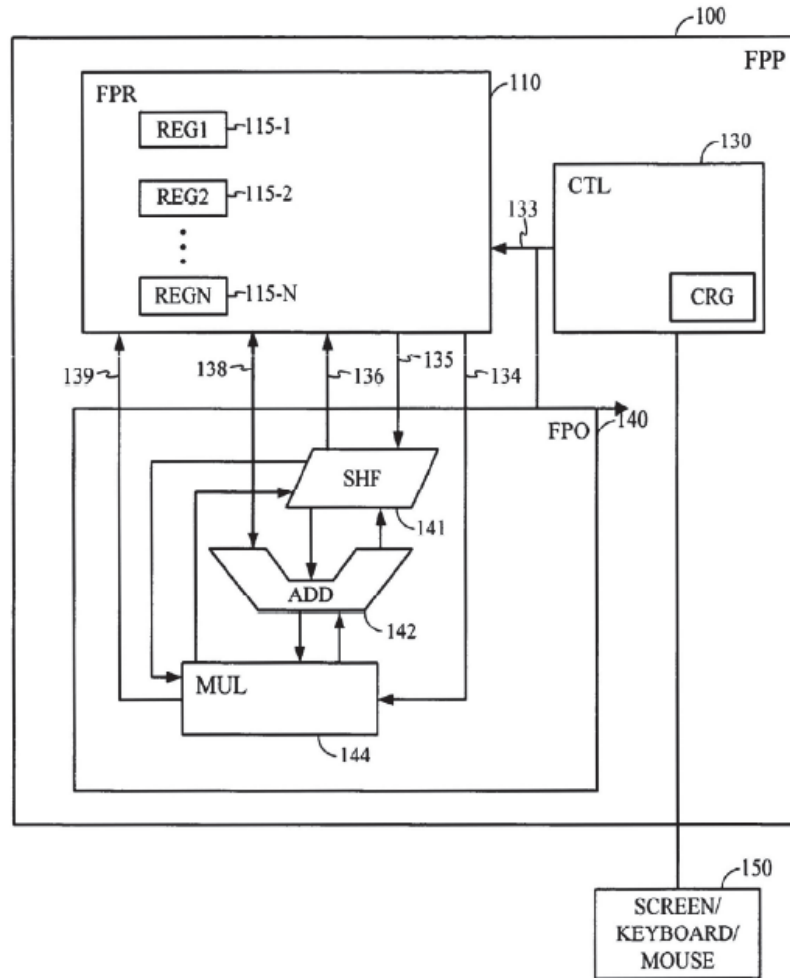
20

IPR2021-00155
Patent 10,416,961 B2

FIG. 1

Figure 1 depicts floating-point processor (FPP) 100 including floating-point register file (FPR) 110 for storing floating-point numbers, floating-point controller (CTL) 130 "used to select the subprecision of the floating-point operations using a control signal 133," and floating-point mathematical operator (FPO) 140 with components "configured to perform the floating-point operations," such as floating-point adder (ADD) 142 and floating-point multiplier (MUL) 144. *Id.* ¶¶ 15, 18, 19.

Figure 2 of Dockser is reproduced below.
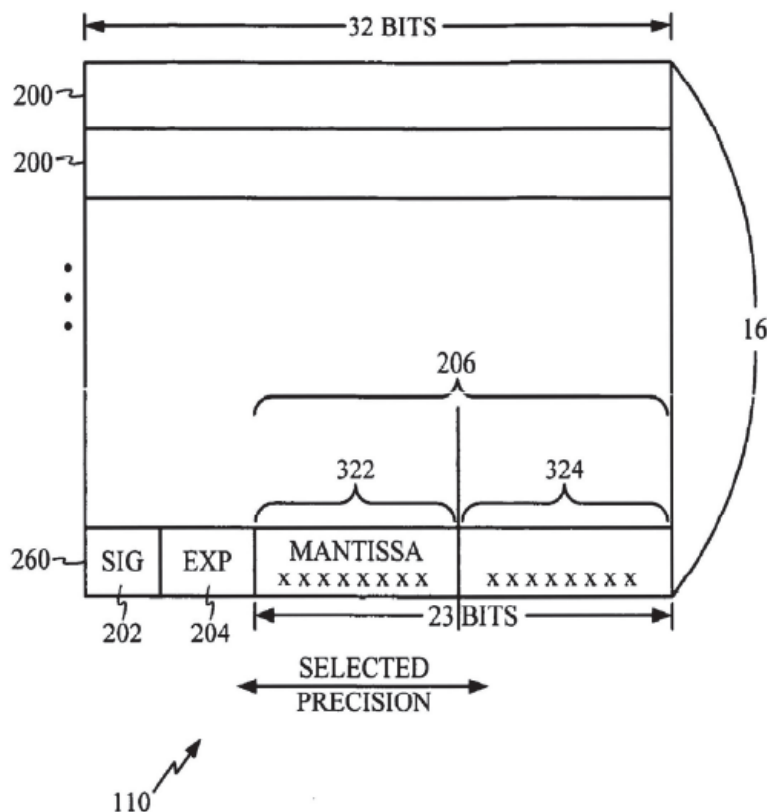
21

IPR2021-00155
Patent 10,416,961 B2

FIG. 2

Figure 2 depicts an exemplary data structure for floating-point register file 110 including 16 addressable register locations 200, each "configured to store a 32-bit binary floating-point number" as "a 1-bit sign 202, an 8-bit exponent 204, and a [23-bit] fraction 206." *Id.* ¶ 17.

"[F]or each instruction of a requested floating-point operation, the relevant computational unit . . . receive[s] from the floating-point register file 110 one or more operands stored in one or more of the register locations" and executes the instruction "at the subprecision selected by the floating-point controller 130." *Id.* ¶¶ 23–24. The precision of the floating-point operation can be reduced by "caus[ing] power to be removed from the floating-point register elements for the excess bits of the fraction that are not required to meet the precision specified by the subprecision select bits" written to the control register. *Id.* ¶¶ 6, 25–26. For example, "if

22

IPR2021-00155
Patent 10,416,961 B2

each location in the floating-point register file contains a 23-bit fraction, and the subprecision required for the floating-point operation is 10-bits, only the 9 commonly significant bits (MSBs) of the fraction are required; the hidden or integer bit makes the tenth." *Id.* ¶ 26. "Power can be removed from the floating-point register elements for the remaining 14 fraction bits." *Id.*

Alternatively, power can be removed in elements of "the logic in the floating-point operator 140 that remains unused as a result of the subprecision selected." *Id.* ¶¶ 7, 27, 29, 32, Fig. 2 (depicting mantissa fraction 206 as having portion 322 for powered bits and portion 324 for unpowered bits). Figures 3A and 3B of Dockser show such removal of power to the floating-point operator logic for a floating-point addition and floating-point multiplication operation, respectively. Figure 3B of Dockser is reproduced below.
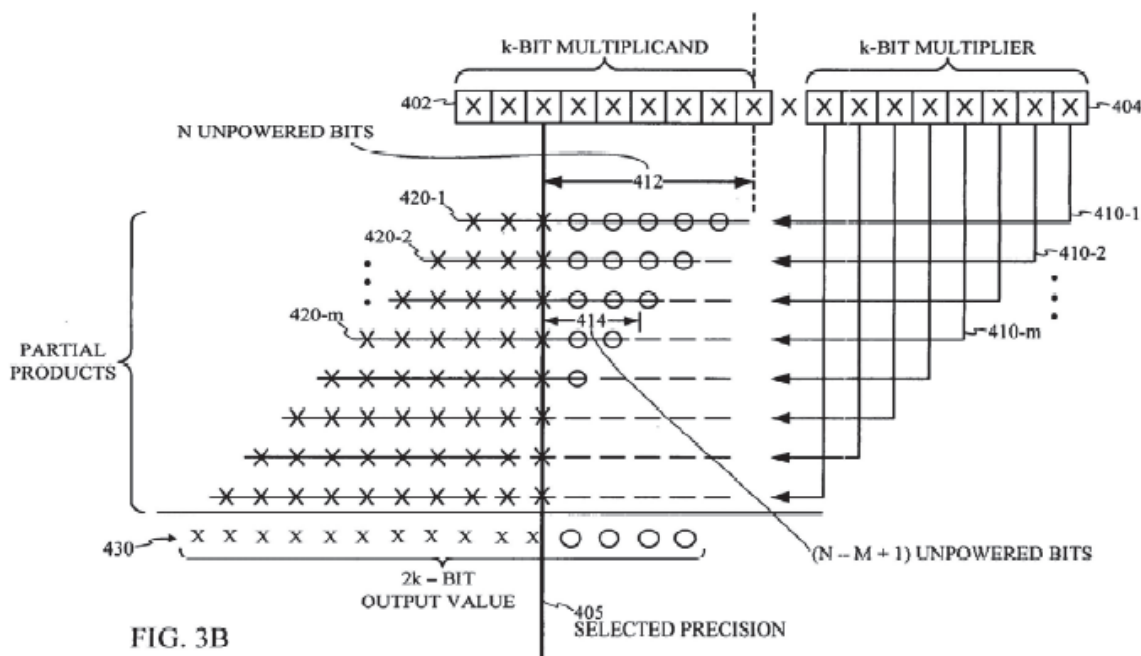

FIG. 3B

Figure 3B depicts k-bit multiplicand 402 and k-bit multiplier 404 to be multiplied together "using a shift-and-add technique," where the multiplication occurs in stages 410-1 through 410-m. *Id.* ¶¶ 30–31.

23

IPR2021-00155
Patent 10,416,961 B2

A partial product 420-i is generated for every bit in multiplier 404 at corresponding stage 410-i and then left-shifted "as a function of the multiplier bit with which it is associated, after which the operation moves on to the next stage." *Id.* ¶ 31. The partial products are eventually added together to generate output value 430. *Id.* "[P]ower may be removed from the logic used to implement the stages to the right of the line 405" indicating the selected subprecision. *Id.* ¶¶ 32–33.

2. *Claim 1*

Petitioner explains in detail how Dockser teaches or suggests every limitation of claim 1, relying on the testimony of Mr. Goodin as support. *See* Pet. 10–39; Goodin Decl. ¶¶ 183–200.

Claim 1 recites, in part, "at least one first low precision high dynamic range (LPHDR) execution unit adapted to execute a first operation on a first input signal representing a first numerical value to produce a first output signal representing a second numerical value." Ex. 1001, 30:18–22.

Petitioner asserts that Dockser's floating-point processor (FPP) is a "low precision high dynamic range (LPHDR) execution unit."[3] Pet. 13–15. According to the Petition, Dockser's FPP is in a computing system "as part of the main processor or as a coprocessor." *Id.* at 13 (citing Ex. 1007 ¶¶ 1, 15, 35–36). Petitioner asserts that the computing system is the claimed

---

[3] Petitioner also provides an "alternative mapping" where "the floating-point operator (FPO) inside Dockser's FPP" constitutes an LPHDR execution unit. Pet. 14–15, 19, 38. We need not evaluate those arguments because we determine that Petitioner has made a sufficient showing that the FPP is an LPHDR execution unit.

IPR2021-00155
Patent 10,416,961 B2

"device" comprising the FPP. *Id.* (citing Goodin Decl. ¶¶ 196–197). We agree. Our reasoning follows.

a)   *"high dynamic range"*

With respect to the "high dynamic range" aspect of the LPHDR execution unit, the claim requires that "the dynamic range of the possible valid inputs to the first operation is at least as wide as from 1/65,000 through 65,000." Ex. 1001, 30:23–25.

Petitioner asserts that Dockser's FPP "uses an 8-bit floating-point exponent . . . that provides an even higher dynamic range" than the 6-bit floating-point exponent disclosed in the '961 patent. Pet. 15 (citing Ex. 1007 ¶ 17; Ex. 1001, 14:61–15:3). Petitioner's assertion, which is not disputed by Patent Owner, is adequately supported by the record.

For example, Dockser's register location 200 stores a 32-bit binary floating-point number, in an IEEE-754 32-bit single format, with 8-bit exponent 204. Ex. 1007 ¶¶ 2, 17, *cited in* Pet. 20. According to the Goodin Declaration, the dynamic range of normal IEEE-754 operands would be from around 2^{-126}, which is much smaller than 1/65,000, to around 2^{127}, which is much larger than 65,000. Goodin Decl. ¶ 232, *cited in* Pet. 20. We credit the Goodin Declaration on this point because it is consistent with Dockser. *See id.*; Ex. 1007 ¶¶ 2, 16–17, 24. Thus, Petitioner has shown Dockser's FPP meets the dynamic-range limitation recited in claim 1.

b)   *"a first operation"*

Claim 1 recites that the LPHDR execution unit is "adapted to execute a first operation on a first input signal representing a first numerical value to

25

IPR2021-00155
Patent 10,416,961 B2

produce a first output signal representing a second numerical value."
Ex. 1001, 30:18–22.

Petitioner argues that Dockser's FPP is adapted to execute a "first operation" (e.g., "reduced-precision multiplication") on a "first input signal representing a first numerical value" (i.e., input electrical signal representing an operand received at the registers) to produce a "first output signal representing a second numerical value" (i.e., output electrical signal representing an operand sent to a register and then main memory). Pet. 19.

Patent Owner does not present an argument to specifically address this argument and evidence. *See* PO Resp.; Sur-Reply.

Based on our review of the totality of the evidence in this case, we agree that Dockser's FPP is adapted to execute the recited first operation. Dockser's FPP receives the operands as input values. Ex. 1007, Fig. 1. Dockser's FPP performs operations on the inputs via the FPP's data paths 134–139 and components 140–144. *Id.*, Fig. 1. We agree with Petitioner that Dockser teaches that the operand data to the FPP is a "first input signal representing a first numerical value" because it represents the operand number. *Id.* ¶ 17. The FPP executes the operation and produces an output value, or output number, which is represented by output bits. *Id.* ¶ 34. In this way, that value is the recited "second numerical value." *See* Pet. 18. We credit the Goodin Declaration on this point. Goodin ¶¶ 223-226.

We also credit the Goodin Declaration's analysis showing that Dockser's computing circuits "move" data between components, such as from memory to processor, via electrical signals. *See* Goodin Decl. ¶ 220 (citing Ex. 1014, col. 1, ll. 36–37). On this issue, the record provides adequate support for the Goodin Declaration. *See* Ex. 1007 ¶¶ 16–17; Ex.

IPR2021-00155
Patent 10,416,961 B2

1014, col. 1, ll. 36–37. Thus, Petitioner has shown that signals represent the recited values.

Considering all the arguments and evidence in this case, Petitioner has sufficiently shown that Dockser teaches that the LPHDR execution unit is "adapted to execute a first operation on a first input signal representing a first numerical value to produce a first output signal representing a second numerical value," as recited in claim 1. Pet. 15–20.

c)   "low precision"

As for the "low precision" aspect of the LPHDR execution unit, Patent Owner argues that Dockser's FPP is not an LPHDR execution unit because the FPP is capable of both low and full precision operations. *See* PO. Resp. 15, 21–24. We disagree with Patent Owner's argument and underlying construction for the reasons discussed in Section II.B. Instead, we agree with Petitioner that (1) the recited "low precision high dynamic range (LPHDR) execution unit" does not exclude units that perform operations at low and full precision, and (2) the degree of "low precision" required by the claim is in limitation *1B2*. *See, e.g.*, Pet. 15, 32; Reply 2–3.

d)   Limitation 1B2

Under the interpretation discussed in Section II.B, Petitioner has shown that Dockser teaches the recited LPHDR execution unit. In particular, Petitioner argues that the FPP is "low precision" because "'the precision' of operations in the FPP is 'reduced'" and because the FPP "operates with the minimum imprecision" required by limitation *IB2*. Pet. 15 (quoting Ex. 1007 ¶ 14). Limitation *1B2* recites,

27

IPR2021-00155
Patent 10,416,961 B2

> for at least X=10% of the possible valid inputs to the first
> operation, the *statistical mean*, over repeated execution of the
> first operation on each specific input from the at least X % of the
> possible valid inputs to the first operation, of the numerical
> values represented by the first output signal of the LPHDR unit
> executing the first operation on that input differs by at least
> Y=0.2% from the result of an *exact mathematical calculation* of
> the first operation on the numerical values of that same input.

Ex. 1001, col. 30, ll. 24–34 (emphasis added). We first analyze the

"statistical mean" limitation before turning to the "exact mathematical

calculation" limitation.

(1) "Statistical Mean"

Petitioner argues that a person of ordinary skill in the art would have

understood the "statistical mean" limitation

> in the context of the '961 patent's stated intent to claim not only
> "repeatable" deterministic embodiments like digital circuits that
> always produce the same output when repeating an operation on
> the same input, but also analog embodiments that are
> non-deterministic because they "introduce noise into their
> computations, so the computations are not repeatable."

Pet. 22 (quoting Ex. 1001, 3:63–4:2).

In Petitioner's view, for "non-deterministic embodiments," the

statistical mean would be the average of "the different outputs produced by

the same operation on the same input." *Id.* Petitioner argues that, for

"deterministic digital embodiments" like Dockser, the statistical mean is

"the same as the numerical value of the first output signal for any individual

execution of the first operation on each specific input, because that output is

always the same for any specific input." *Id.* at 23. In Petitioner's view,

repeatedly executing a multiplication operation using Dockser's floating-

point multiplier "on the same input (*i.e.*, pair of operands) with the same

IPR2021-00155
Patent 10,416,961 B2

precision level yields the same result for every execution; therefore, the statistical mean of the outputs is the same as the output for any single execution." *Id.*

Patent Owner does not dispute this reasoning. *See* PO Resp.; Sur-Reply.

Based on the totality of the evidence, Petitioner's analysis of "deterministic digital embodiments" to address the "statistical mean" limitation is adequately supported by the record. Pet. 22–23. We are persuaded by Petitioner's reasoning that the '961 patent's written description has both analog and digital embodiments. *See, e.g.*, Ex. 1001, col. 3, l. 63–col. 4, l. 2; col. 14, l. 23–col. 15, l. 3; col. 17, ll. 22–26. Dockser uses "a conventional floating-point multiplier." Ex. 1007 ¶ 20. Petitioner adequately supports its assertion that this is a deterministic digital circuit. Pet. 23 (citing Goodin Decl. ¶ 241; Ex. 1011, col. 1, ll. 40–42). And we credit the Goodin Declaration that, in such a digital circuit, a floating-point multiplier operating on the same input with the same precision level yields the same result for every execution, and thus, the statistical mean is the same as the output for any single execution. Goodin Decl. ¶ 240, *cited in* Pet. 23.

(2)   "exact mathematical calculation"

As for the "exact mathematical calculation" limitation, Petitioner argues that because Dockser performs a reduced precision multiplication, the result of the operation differs from what would be the exact mathematical result of the operation: "the (>32-bit) product that would result if the pair of input 32-bit operands were multiplied without reducing precision." Pet. 24 (emphasis omitted) (citing Goodin Decl. ¶ 244; Appendix I.B).

IPR2021-00155
Patent 10,416,961 B2

(3) Y Relative Error for X Valid Inputs

According to the Petition, "Dockser discloses two precision-reducing techniques that can be used separately or together." Pet. 26 (citing Ex. 1007 ¶¶ 4–7; Goodin Decl. ¶¶ 255–261). One technique removes power from storage elements in the FPP registers that correspond to the excess, dropped mantissa bits, which drops bits from the operands. *Id.* The other technique removes elements from the multiplier logic that computes the product of the operand mantissas. *Id.* Petitioner asserts that both techniques individually and in combination teach limitation *1B2*.

We determine that Petitioner has made a sufficient showing that Dockser meets the recited requirements for low precision under the register bit-dropping rationale. Patent Owner does not present arguments specific to this technique. *See* PO Resp.; Sur-Reply. We need not evaluate Petitioner's alternative rationale based on the multiplier logic. Our reasoning follows.

(4) Register Bit-Dropping

Petitioner explains how Dockser teaches that "for at least X=10% of the possible valid inputs to the first operation," the statistical mean of the results of executing the first operation "differs by at least Y=0.2%" from the result of the exact mathematical calculation. Pet. 24–39. In Petitioner's view, the "possible valid inputs" in Dockser are "the set of possible normal IEEE-754 32-bit single-format numbers forming pairs of operands in input signals to the execution unit that can be multiplied together to produce an output representing a numerical value (rather than, *e.g.*, an overflow/underflow exception)." *Id.* at 25. Petitioner argues that Dockser's FPP operates at a precision level meeting the claimed X and Y percentages for such input pairs. *Id.*

IPR2021-00155
Patent 10,416,961 B2

We agree. Dockser retains only some of the bits of a mantissa fraction. In one example, nine most-significant bits of the mantissa fraction are retained, and the other 14 excess bits are dropped. Ex. 1007 ¶¶ 25–28, *cited in* Pet. 25; Goodin Decl. ¶ 254. The outputs of Dockser's unpowered storage elements are tied to zero voltage. Goodin Decl. ¶ 269; Ex. 1007 ¶¶ 26 (discussing removing power from the register elements for the excess bits), 29 (discussing "unpowered bits"). With support from the Goodin Declaration and corroborating evidence, Petitioner shows that tying the outputs of the unpowered storage elements for the 14 "excess" bits to zero voltage would make those bits zeroes. Pet. 29–30 (citing Goodin Decl. ¶¶ 269–71).

Multiplying two operands with excess bits dropped (representing zero) reduces the output's precision. Ex. 1007 ¶¶ 24–26. Petitioner provides a representative example from the Goodin Declaration:

> the 23-bit sequence 01100000111100111001110 (representing a mantissa of 1.3787171840667724609375) would become 01100000100000000000000 (representing a mantissa of 1.376953125) by zeroing the 14 right-most bits, and 01000100110111000001000 (representing a mantissa of 1.26898288726806640625) would become 01000100100000000000000 (representing a mantissa of 1.267578125).

Pet. 30 (citing Goodin Decl. ¶ 273). So the decrease in precision depends on the number of mantissa bits dropped. *See id.*; *see also* Goodin Decl. ¶ 278. And we agree with Petitioner that Dockser suggests that any desired number of mantissa bits can be dropped. Pet. 32. For example, Dockser describes how the precision is selectable. *See* Ex. 1007 ¶¶ 14–15, 17 (describing selectable reduced precision), *discussed in* Pet. 5.

IPR2021-00155
Patent 10,416,961 B2

Petitioner asserts that Dockser's register bit-dropping technique meets limitation *1B2* in two ways. Pet. 27–38. First, Petitioner presents results of a software demonstration to show that Dockser's 9 mantissa bits meets limitation *1B2*. *Id.* at 32. Second, Petitioner uses a pencil-and-paper algebraic analysis to show that retaining 7 mantissa bits meets limitation *1B2*. *Id.* at 33.

As for the software demonstration, Petitioner states that "[g]iven the massive number of possible inputs to Dockser's FPP (including over 70 trillion possible pairs of normal IEEE-754 single-format mantissas), a [person of ordinary skill in the art] would have performed Dockser's FPP operation in software to determine the fraction X of all possible valid inputs that produce at least the claimed relative error Y when a given number of mantissa bits are dropped." *Id.* at 32. Petitioner states that Mr. Goodin wrote a program to perform reduced precision multiplication retaining 9 mantissa bits and dropping the 14 excess bits as in Dockser that tested all possible valid mantissa pairs and "produce[d] at least Y=0.2% relative error for 14.6% of possible valid inputs (greater than X=10%)." *Id.* at 33, 66–68 (citing Goodin Decl. ¶¶ 412–426, 464–469).

Patent Owner did not rebut this argument and evidence. *See* PO Resp.; Sur-Reply.

We are persuaded by Petitioner's assertions about recited X and Y percentages, and we credit the Goodin Declaration on this issue, which is supported by the record. *See* Goodin Decl. ¶¶ 412–426. In particular, Mr. Goodin summarizes the results of the software program in the table below.

IPR2021-00155
Patent 10,416,961 B2

Retained Fraction Bits	Claimed Y%	Percentage of Tested Pairs Meeting Claimed Y%	Adjusted "Dockser's X%"
9	≥ .2%	14.674747%	≥ 14.59791%
5	≥ .2%	99.547206%	≥ 99.02600%

The table shows the percentage of tested pairs that meet the claimed Y percentage as output by the software program. Goodin Decl. ¶ 423. It also shows a "Dockser's X" percentage, which is adjusted to exclude possible overflow and underflow pairs, that meets the claimed requirements. *Id.* The X value is calculated by multiplying the percentage output by the software program by an adjustment ratio. *Id.* The adjustment ratio is described in Appendix I.D of the Goodin Declaration. *Id.* The Goodin Declaration also explains that, if "the 'exact mathematical calculation' of multiplication of two single-format floating-point operands is the product rounded or truncated to single format (having a 23-bit-fraction mantissa), the claims are still met." *Id.* ¶ 426. To show this, Mr. Goodin wrote an alternative version of the program under this assumption, which also gives results that meet the claim limitations. *Id.* After review of the arguments and evidence on these issues, we credit the Goodin Declaration's software-based analysis. *Id.* ¶¶ 412–426.

As for the algebraic analysis, Petitioner argues that a person of ordinary skill in the art "would also have understood algebraically that Dockser's register bit-dropping technique meets [the limitation] when performed at certain selected precision levels, by examining the absolute *minimum* relative error produced by zeroing certain mantissa bit positions." Pet. 33. Petitioner explains how over 12% of possible input operands "have a zero as their most-significant (leftmost) mantissa fraction bit and ones as

33

IPR2021-00155
Patent 10,416,961 B2

their eighth and ninth fraction bits," such that retaining only 7 mantissa bits of operands would result in "*every* input in that 12% produc[ing] at *minimum* 0.39% relative error." *Id.* at 33, 68–72 (citing Goodin Decl. ¶¶ 427–444).

Patent Owner did not rebut this argument and evidence. *See* PO Resp.; Sur-Reply.

We are persuaded by Petitioner's algebraic method showing recited X and Y percentages and credit the Goodin Declaration on this issue. *See* Goodin Decl. ¶¶ 427–444. Under this reasoning, for the 12.5% of all possible M_A mantissa values with zero in the first fraction bit and ones in the $(K + 1)^{th}$ and $(K + 2)^{th}$ fraction bits, the following inequality holds when Dockser's register bit-dropping truncates the operands to K fraction bits:

$$Y \geq \frac{\left(2^{-(K+1)} + 2^{-(K+2)}\right)}{(1.5 - 2^{-23})} \times 100\% \qquad \textbf{\textit{(Equation C)}}$$

Id. ¶ 442. The table below shows the results of evaluating Equation C with K=7 and K=5 retained fraction bits as the selected precision level. *Id.*

Retained Fraction Bits (K)	X%	Y%
7	≥ 12%	≥ 0.3906%
5	≥ 12%	≥ 1.5625%

The table above shows that both values of K meet the '961 patent's claimed minimum X and Y. *Id.* ¶ 444. After review of the arguments and evidence on these issues, we credit the Goodin Declaration's algebraic analysis. *Id.* ¶¶ 427–444.

IPR2021-00155
Patent 10,416,961 B2

e) "control the operation"

Lastly, claim 1 recites, "at least one first computing device adapted to control the operation of the at least one first LPHDR execution unit." Petitioner asserts that Dockser's main processor meets the computing device limitation. Pet. 39.

Patent Owner does not provide evidence or argument specific to this limitation.

We agree with Petitioner because Dockser's processor writes subprecision select bits to the FPP's control register, and in at least this way, Dockser's main processor is adapted to control the FPP's operation by specifying its precision level. *See* Ex. 1007 ¶¶ 15, 18, 25, 35. We credit the Goodin Declaration on this issue, which is supported by the evidence of record. *See* Goodin Decl. ¶ 310.

f) *Objective Indicia of Nonobviousness*

In addition to its argument based on the interpretation of "low precision high dynamic range (LPHDR) execution unit," Patent Owner argues that objective indicia of nonobviousness demonstrate that the challenged claims would not have been obvious to a person of ordinary skill in the art. PO Resp. 47–72. Because the parties largely refer to the challenged claims collectively in their arguments regarding objective indicia of nonobviousness, we do so as well and now address those arguments as applied to all of the claims challenged as obvious over Dockser alone or the combination of Dockser and Tong, i.e., claims 1, 2, 4, 5, 10, 13, 14, 21, 24,

IPR2021-00155
Patent 10,416,961 B2

and 25, which we refer to as the "Dockser/Tong-challenged claims."[4] *See id.*; Reply 19–30; Sur-Reply 18–25.

"In order to accord substantial weight to [objective indicia] in an obviousness analysis, the evidence of [objective indicia] must have a nexus to the claims, *i.e.*, there must be a legally and factually sufficient connection between the evidence and the patented invention." *Fox Factory, Inc. v. SRAM, LLC*, 944 F.3d 1366, 1373 (Fed. Cir. 2019) (citations and internal quotation marks omitted). "The patentee bears the burden of showing that a nexus exists." *Id.* "To determine whether the patentee has met that burden, we consider the correspondence between the objective evidence and the claim scope." *Id.* A patentee is entitled to a rebuttable presumption of nexus "when the patentee shows that the asserted objective evidence is tied to a specific product and that product 'embodies the claimed features, and is coextensive with them.'" *Id.* (citation omitted). Whether a rebuttable presumption of nexus arises "turns on the nature of the claims and the specific facts." *Teva Pharms. Int'l GmbH v. Eli Lilly & Co.*, 8 F.4th 1349, 1362 (Fed. Cir. 2021). "As part of the presumption analysis, the fact finder must consider the unclaimed features of the stated products to determine their level of significance and their impact on the correspondence between

---

[4] As explained below, we determine that Petitioner has not proven that certain claims are unpatentable over the combinations of Dockser and MacMillan, and Dockser, Tong, and MacMillan, because Petitioner has not sufficiently shown a reason to combine with MacMillan. *See infra* Sections II.F, II.G. Accordingly, we need not assess objective indicia of nonobviousness with respect to those claims. *See Mylan Pharms. Inc. v. Res. Corp. Techs., Inc.*, 914 F.3d 1366, 1376 (Fed. Cir. 2019) ("Because we agree with the Board that Appellants failed to establish a motivation to modify compound 3I, we need not reach Appellants' arguments regarding objective indicia.").

IPR2021-00155
Patent 10,416,961 B2

the claim and the products." *Quanergy Sys., Inc. v. Velodyne Lidar USA, Inc.*, 24 F.4th 1406, 1418 (Fed. Cir. 2022). When, "for example, the patented invention is only a small component of the product tied to the objective evidence, there is no presumption of nexus." *Henny Penny Corp. v. Frymaster LLC*, 938 F.3d 1324, 1333 (Fed. Cir. 2019). Absent a presumption of nexus, a patent owner may "prove nexus by showing that the evidence of secondary considerations is the 'direct result of the unique characteristics of the claimed invention.'" *Fox Factory*, 944 F.3d at 1373–74 (citation omitted).

"Where the offered secondary consideration actually results from something other than what is both claimed and *novel* in the claim, there is no nexus to the merits of the claimed invention," meaning that "there must be a nexus to some aspect of the claim not already in the prior art." *In re Kao*, 639 F.3d 1057, 1068–69 (Fed. Cir. 2011). On the other hand, there is no requirement that "objective evidence must be tied exclusively to claim elements that are not disclosed in a particular prior art reference in order for that evidence to carry substantial weight." *WBIP, LLC v. Kohler Co.*, 829 F.3d 1317, 1331 (Fed. Cir. 2016). A patent owner may show, for example, "that it is the claimed combination as a whole that serves as a nexus for the objective evidence; proof of nexus is not limited to only when objective evidence is tied to the supposedly 'new' feature(s)." *Id.* at 1330. Ultimately, the fact finder must weigh the objective indicia of nonobviousness presented in the context of whether the claimed invention as a whole would have been obvious to an ordinarily skilled artisan. *Id.* at 1331–32; *Lectrosonics, Inc. v Zaxcom, Inc.*, IPR2018-01129, Paper 33 at 33 (PTAB Jan. 24, 2020) (precedential).

37

IPR2021-00155
Patent 10,416,961 B2

(1) Industry Skepticism

"If industry participants or skilled artisans are skeptical about whether
or how a problem could be solved or the workability of the claimed solution,
it favors [nonobviousness]. Doubt or disbelief by skilled artisans regarding
the likely success of a combination or solution weighs against the notion that
one would combine elements in references to achieve the claimed
invention." *WBIP*, 829 F.3d at 1335. As evidence of alleged industry
skepticism, Patent Owner points to various statements made by employees
of Petitioner after meeting with Dr. Bates, the inventor of the '961 patent, in
2010. PO Resp. 48–52. According to Patent Owner, Dr. Bates explained
that floating point units programmed to perform ███████████

███

███████████████████████████████████

██

██████████████████████████████████

██████████████████████ *Id.* at 49–52 (citing Exs. 2009, 2027, 2028,
2030). Patent Owner contends that the cited statements from Petitioner's
employees "express skepticism about utility of the low-precision arithmetic
performed by the large numbers of low-precision execution units,"[5] which is
a "key feature" of the challenged claims, and have a nexus to the challenged

---

[5] Claim 1 recites "at least one" LPHDR execution unit and thus is satisfied
by only a single LPHDR execution unit. The only challenged claims of the
'961 patent that arguably require a "large number[]" of LPHDR execution
units are challenged in Petitioner's asserted grounds based on combinations
with MacMillan. *See* PO Resp. 52. For example, claims 3 and 23 recite "the
number of LPHDR execution units in the device exceeds by at least one
hundred the non-negative integer number of execution units in the device
adapted to . . . ." Ex. 1001, col. 30, ll. 43–48 (claim 3); col. 32, ll. 62–65
(claim 23).

IPR2021-00155
Patent 10,416,961 B2

claims because they were made in response to Dr. Bates's presentations. *Id.* at 52.

Petitioner argues that Patent Owner's evidence is irrelevant because it is dated after the earliest possible effective filing date of the '961 patent (June 19, 2009). Reply 22. We disagree that evidence of industry skepticism must itself be dated before the relevant invention. *See WBIP*, 829 F.3d at 1335–36 (considering evidence of skepticism dated after the invention of the patent at issue); *Pressure Prods. Med. Supplies, Inc. v. Greatbatch Ltd.*, 599 F.3d 1308, 1319 (Fed. Cir. 2010) (same); Sur-Reply 19–20.

We have reviewed the substance of the statements cited by Patent Owner, though, and are not persuaded that they are probative of nonobviousness. The statements by Petitioner's engineers relate to whether applications using "approximate" arithmetic would be commercially valuable; they do not show skepticism as to whether or how a problem could be solved or whether such applications would *work*. *See WBIP*, 829 F.3d at 1335. For example, Tom Dean stated that

Ex. 2009, 1 (emphasis added).

Ex. 2029, 1

(emphasis added).

39

IPR2021-00155
Patent 10,416,961 B2

Ex. 2028, 1 (emphasis added).

Ex. 2030, 2. Thus, the cited comments appear to indicate that floating point units performing "approximate" arithmetic would only be commercially valuable for a limited set of applications, rather than expressing skepticism as to whether such applications would work. Patent Owner also has not explained sufficiently why the fact that "approximate" arithmetic was considered useful for some applications but not others, amounts to industry skepticism over the claimed invention as a whole, which is not limited to a particular application. We find that Patent Owner's evidence of alleged industry skepticism does not weigh in favor of nonobviousness of the claims challenged under the Dockser-Tong combination.

(2) Industry Praise

"Evidence that the industry praised a claimed invention or a product which embodies the patent claims weighs against an assertion that the same claim would have been obvious. Industry participants, especially competitors, are not likely to praise an obvious advance over the known art." *WBIP*, 829 F.3d at 1334. Patent Owner points to four statements of alleged industry praise from "leading figures in computer science" and argues that they have a nexus to the Dockser/Tong-challenged claims because they relate to Dr. Bates and his presentations to Petitioner in 2010, 2011, and 2013. PO Resp. 52–53.

IPR2021-00155
Patent 10,416,961 B2

Patent Owner cites Dr. Bates's notes from ▮▮▮▮▮▮▮▮▮▮▮▮▮

▮▮▮

▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮ *See id.* at 52; Ex. 2050. This single

line in the notes does not mention any specific ideas or features that

allegedly were praised, or why or for what they were being praised. Patent

Owner argues that there is "no requirement that praise identify exactly the

features being praised." Sur-Reply 20. But evidence of industry praise must

be for "the claimed invention," and the notes only state that ▮▮▮▮▮▮▮

▮▮▮▮▮▮▮▮▮▮▮▮▮ *See* Ex. 2050; *Apple Inc. v. Samsung Elecs. Co.,*

839 F.3d 1034, 1053 (Fed. Cir. 2016) (en banc); *WBIP,* 829 F.3d at 1334.

The notes also have less probative value because they were written by

Dr. Bates based on his subjective recollection of Mr. Dean's reaction at the

meeting, rather than, for example, being prepared by a third party or Mr.

Dean himself. Ex. 2052 ¶ 7.

Patent Owner also cites three 2011 emails sent by employees of

Petitioner to Dr. Bates. PO Resp. 52–53. Andrew Ng wrote the following to

Dr. Bates:

▮▮

Ex. 2033, 1. Astro Teller wrote that ▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮

▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮ and Johnny Chen wrote, ▮▮▮▮▮▮▮▮

▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮ *See*

IPR2021-00155
Patent 10,416,961 B2

Exs. 2031, 2032. We view these emails as more consistent with Petitioner's view as "merely cordial statements made to Dr. Bates," rather than evidence of industry praising the inventions of the Dockser/Tong-challenged claims as Patent Owner contends. *See* Reply 24. Further, they are statements by employees of Petitioner only (not any other industry participant or industry group) and do not mention any specific ideas or features from Dr. Bates's presentations that were being praised. We find that Patent Owner's evidence of alleged industry praise is too ambiguous and high-level to be probative of nonobviousness.

(3) Unexpected Results

"To be particularly probative, evidence of unexpected results must establish that there is a difference between the results obtained and those of the closest prior art, and that the difference would not have been expected by one of ordinary skill in the art at the time of the invention." *Bristol-Meyers Squibb Co. v. Teva Pharms. USA, Inc.*, 752 F.3d 967, 977 (Fed. Cir. 2014). Also, "[w]hen assessing unexpected properties, . . . we must evaluate the significance and 'kind' of expected results along with the unexpected results." *Id.* As evidence of alleged unexpected results, Patent Owner points to a September 2013 presentation that Dr. Bates gave to Petitioner showing

██

████████████████████ PO Resp. 53–57 (citing Ex. 2005). Patent Owner argues that the unexpected nature of the results is "confirmed by skepticism expressed by" David Patterson, Ph.D. (an employee of Petitioner and "acclaimed expert in the field of computer science and computer architecture" according to Patent Owner), and "further supported by Dockser

42

IPR2021-00155
Patent 10,416,961 B2

and Tong," which teach that higher precision is "generally" needed. *Id.* at 56–57 (citing Ex. 1003 ¶¶ 124–125; Ex. 1007 ¶ 3; Ex. 1008, 280; Ex. 2012).

 The cited evidence, however, does not establish a difference between the alleged results and the closest prior art. Dr. Bates's presentation includes ██████████████████████████████████ Ex. 2005, 11; *see* PO Resp. 55. In this case, however, the closest prior art is Dockser and Tong, both of which teach reduced precision arithmetic and, importantly, show that high performance with reduced precision was *expected* for certain applications. *See* Reply 25; Ex. 1007 ¶ 3 ("[F]or certain applications, . . . a reduced precision, may be acceptable, and for other applications, . . . a greater precision may be needed."); Ex. 1008, 273 ("[M]any programs which manipulate human sensory inputs, e.g., speech and image recognition, suffer no loss of accuracy with reduced bitwidth in the mantissa or exponent.")." Patent Owner contends that Dockser and Tong are not appropriate for comparison because they do not "provide computing scale results" and have an "entirely different focus on power savings." Sur-Reply 21. We disagree. As explained above, we find that Dockser teaches the claim 1 LPHDR execution unit performing the recited reduced precision arithmetic. *See supra* Sections II.D.2.a)–e). Patent Owner's only comparison, however, is to a CPU without any reduced precision capability. *See* PO Resp. 55. We find that Patent Owner's evidence of alleged unexpected results does not weigh in favor of nonobviousness of the Dockser/Tong-challenged claims because it does not establish a difference between the results and the closest prior art and does not establish that such a difference would have been unexpected by a person of ordinary skill in the art.

IPR2021-00155
Patent 10,416,961 B2

(4) Copying and Commercial Success

"Copying may . . . be another form of flattering praise for inventive features," as "[t]he fact that a competitor copied technology suggests it would not have been obvious." *WBIP*, 829 F.3d at 1336 (citation omitted). Copying "requires evidence of efforts to replicate a specific product." *Wyers v. Master Lock Co.*, 616 F.3d 1231, 1246 (Fed. Cir. 2010). "This may be demonstrated either through internal documents; direct evidence such as disassembling a patented prototype, photographing its features, and using the photograph as a blueprint to build a virtually identical replica; or access to, and substantial similarity to, the patented product (as opposed to the patent)." *Iron Grip Barbell Co. v. USA Sports, Inc.*, 392 F.3d 1317, 1325 (Fed. Cir. 2004) (internal citations omitted). However, "a showing of copying is only equivocal evidence of non-obviousness in the absence of more compelling objective indicia of other secondary considerations." *Ecolochem, Inc. v. S. Cal. Edison Co.*, 227 F.3d 1361, 1380 (Fed. Cir. 2000); *see also In re GPAC Inc.*, 57 F.3d 1573, 1580 (Fed. Cir. 1995) ("[M]ore than the mere fact of copying by an accused infringer is needed to make that action significant to a determination of the obviousness issue." (citation omitted)).

Commercial success is typically shown with evidence of "significant sales in a relevant market." *Ormco Corp. v. Align Tech., Inc.*, 463 F.3d 1299, 1312 (Fed. Cir. 2006) (citation omitted). "When a patentee can demonstrate commercial success, usually shown by significant sales in a relevant market, and that the successful product is the invention disclosed and claimed in the patent, it is presumed that the commercial success is due to the patented invention." *J.T. Eaton & Co. v. Atlantic Paste & Glue Co.*, 106 F.3d 1563, 1571 (Fed. Cir. 1997).

IPR2021-00155
Patent 10,416,961 B2

Patent Owner argues that after Dr. Bates made his presentations to
Petitioner in 2010–2014, Petitioner copied Dr. Bates's invention in its
Tensor Processing Unit (TPU) v2 and v3 products released in 2017–2018,
which "use execution units designed to perform operations using a
low-precision 'floating point format called bfloat16' in order to achieve
increased parallelism and scale." PO Resp. 57–58 (citing Ex. 2010; Ex.
2011, 9; Ex. 2034, 6; Khatri Decl. ¶¶ 126–127). The bfloat16 format
"keep[s] the same 8 bit exponent as found in 32-bit IEEE floating point" but
"reduces precision significantly by using only a 7-bit mantissa, as opposed
to the 23-bit mantissa of IEEE 32-bit floating point." *Id.* at 58 (citing Ex.
2041, 2). According to Patent Owner, Petitioner uses the TPUv2 and TPUv3
products to "power all of its major products and services, including Search,
Translate, Photos, Assistant, and Gmail," saving Petitioner "at least $10
billion." *Id.* at 61–62 (citing Exs. 2011, 2015–18, 2040). Patent Owner
argues that a nexus is presumed because the TPUv2 and TPUv3 products are
coextensive with claims 1–5, 10, and 13–14. *Id.* at 59, 62–72. Patent Owner
further contends that, even if not presumed, a nexus is shown by "internal
communications among [Petitioner's] own engineers and executives" and
Petitioner's communications with Dr. Bates. *Id.* at 59–61 (citing Exs. 2007,
2014, 2035–2039, 2042, 2047, 2048).

We have reviewed the cited evidence and do not find it probative of
nonobviousness. First, we are not persuaded that a nexus should be
presumed because the TPU products are coextensive with the Dockser/Tong-
challenged claims, as Patent Owner contends. Petitioner and Mr. Goodin
identify a number of other allegedly material features of the products that are
not part of the Dockser/Tong-challenged claims: "Inter-Core Interconnect"
(ICI), "Systolic array," "Two TensorCores per chip," "High-bandwidth

IPR2021-00155
Patent 10,416,961 B2

memory," "Liquid cooling," and "Accelerated Linear Algebra compiler."

See Reply 26; Ex. 1071 ¶¶ 85–92. We focus on the first two features in

particular because the evidence of record characterizes them as significant.

Regarding the ICI of the TPU products, Dr. Patterson stated the

following:

> So the *critical feature* of the supercomputer is the way
> they talk to each other – you know, the chips talk to each other.
> So what's called – [Petitioner] invented something called ICI.
> It's a custom Inter-Core Interconnect that they did custom for
> this design; it's not some standard that you buy off the shelf.
> It runs at about 500 gigabits per section in both directions per
> link, and there's four of those links.

Ex. 2012, 14:17–24 (emphasis added). An article authored by Dr. Patterson

and others[6] states that "[t]he *critical architecture feature* of a modern

supercomputer is how its chips communicate: what is the speed of a link;

what is the interconnect topology; does it have centralized versus distributed

switches; and so on," and the TPUv2 product has "four custom Inter-Core

Interconnect (ICI) links, each running at 496Gbits/s per direction."

Ex. 2016, 69; *see id.* at 70, 72, Fig. 2 (showing that ICI is part of the TPUv3

product as well); Ex. 1071 ¶ 87.

Regarding the systolic array architecture of the TPU products,

Dr. Patterson stated as follows:

> This systolic array idea brought back from the past that
> probably *cut the energy costs in half* for the multiplier, which is
> a lot of it. And then it was domain-specific. So a lot of the things
> that are complicated and area consuming on CPUs like branch

---

[6] We disagree with Patent Owner that the materials cited herein are merely
"marketing documents." *See* Conf. Tr. 13:18–24. Dr. Patterson's comments
are from an audio transcription of a YouTube video filed by Patent Owner
(Ex. 2012), and the article is from the publication *Communications of the
ACM* (Ex. 2016). *See* PO Resp. ix.

IPR2021-00155
Patent 10,416,961 B2

> predictors and caches – CPUs are covered with caches – they're not there because they're not helpful for this application. It was better to use more multiple units and dedicated memory units. So that *saves energy* and you could reuse the resources. So that was a *big success*.

Ex. 2012, 9:9–18 (emphasis added). He also agreed with an article comparing the TPU to graphics processing units (GPUs) stating that "GPUs incur high overhead in performance, area, and energy due to heavy multithreading," whereas "[t]he systolic organization of TPUs captures the data reuse while being simple by avoiding multithreading." *Id.* at 33:20–34:2. The article authored by Dr. Patterson and others includes similar statements. Ex. 2016, 70 ("The TPUv2 node of the supercomputer followed the main ideas of TPUv1: A large two-dimensional matrix multiply unit (MXU) *using a systolic array to reduce area and energy* plus large, software-controlled on-chip memories instead of caches." (emphasis added)), 74 (quoting the earlier article stating that the "systolic organization" of the TPU captures data reuse "while being simple by avoiding multithreading"); *see* Ex. 1071 ¶ 88.

Patent Owner contends that Petitioner fails to show that the "features are so critical that the TPUs are not 'essentially the invention.'" Sur-Reply 21. We are not persuaded. Petitioner does more than merely identify unclaimed features of the TPU products. Petitioner cites evidence, supported by the testimony of Mr. Goodin, showing that at least the ICI and systolic array architecture materially affect the TPU products' performance and were viewed as critical, successful features that were reasons for the TPU products' improved performance. *See Fox Factory*, 944 F.3d at 1374 ("[I]f the unclaimed features amount to nothing more than additional insignificant features, presuming nexus may nevertheless be appropriate."),

47

IPR2021-00155
Patent 10,416,961 B2

1375–76 (noting that the challenged patent and the patent owner's "marketing materials confirm that the forwardly protruding tooth tips, hook features, and mud clearing recesses each materially impacts the functioning of a chainring," such that "[f]or each of these features that the Board confirms is included in the [allegedly coextensive] chainrings, nexus can only be presumed between the . . . chainrings and a patent claim if the claim includes limitations relating to these features"). Patent Owner does not point to any evidence indicating that view to be factually incorrect. *See* Ex. 1072, 147:15–149:2 (Dr. Khatri opining that "the TPUs have the features that . . . are in these claims" and stating that he has no opinion "about what the TPUs . . . might have in addition"). We find that a presumption of nexus is inappropriate with respect to the TPU products.

Regardless of whether a presumption of nexus applies, we are not persuaded that Patent Owner has made a sufficient showing of nexus for two reasons. First, Patent Owner's nexus arguments are premised on the TPUv2 and TPUv3 products' "low-precision" bfloat16 floating point format having an 8-bit exponent and 7-bit mantissa. PO Resp. 57–58 (arguing that Petitioner copied Dr. Bates's invention by implementing "execution units designed to perform operations using" the bfloat16 format), 62 (arguing commercial success due to the bfloat16 format using "low-precision multipliers"). But "[l]ow-precision arithmetic—including LPHDR multiplication—was not novel" and thus cannot be the basis for a finding of nexus. *See* Reply 20; *Kao*, 639 F.3d at 1068; *Tokai Corp. v. Easton Enterprises, Inc.*, 632 F.3d 1358, 1369 (Fed. Cir. 2011) ("If commercial success is due to an element in the prior art, no nexus exists."). Tong taught such a format nine years before the '961 patent. Tong describes two different experiments emulating "different bitwidth FP units": one that

IPR2021-00155
Patent 10,416,961 B2

"implements the IEEE-754 standard" and then varies the mantissa bitwidth (shown in Figure 6), and another that "implements the IEEE-754 standard" and then varies the exponent bitwidth (shown in Figure 7). Ex. 1008, 278–279. Thus, we agree with Mr. Goodin that a person of ordinary skill in the art "would have understood that Tong's Figure 6 . . . discloses a floating-point format with an 8-bit exponent (which is the exponent size of single-precision floating-point in the IEEE standard) . . . and a 7-bit mantissa." *See* Ex. 1071 ¶ 74 (citing Ex. 1008, 274, 278–279; Ex. 1003 ¶ 330). Tong discloses that the 8-bit exponent/7-bit mantissa format produced useful results for both a "neural network trainer" application (ALVINN) and a "speech recognition program" (Sphinx III). *See* Ex. 1008, 278–279; Ex. 1071 ¶ 75. Notably, these applications mirror Dr. Bates's 2013 presentation discussing ███████████████████████████" *See* Ex. 2005, 12–13.

To be sure, the fact that the 8-bit exponent/7-bit mantissa format was known in the prior art is not dispositive. *See WBIP*, 829 F.3d at 1330–31. "Commercial success, for example, may be linked to an individual element or, in other circumstances, it could be linked to the inventive combination of known elements." *Id.* at 1332. Patent Owner, however, has not identified and explained a specific combination of elements (as opposed to just the "low-precision" bfloat16 floating point format) that could serve as a nexus for the evidence of alleged copying and commercial success.[7] *See* PO Resp. 57–58, 61–62.

---

[7] To the extent Patent Owner alleges commercial success from Petitioner "[u]sing hundreds of thousands of these low-precision TPV v2 and v3 devices," we note again that claim 1 recites "at least one" LPHDR execution unit and thus is satisfied by a single unit. *See* PO Resp. 61.

IPR2021-00155
Patent 10,416,961 B2

Second, when arguing that the TPUv2 and TPUv3 products meet the X and Y percentage limitations of claim 1, Patent Owner points to the bfloat16 floating point format with a 7-bit mantissa and states that Petitioner argues in the Petition (with respect to Dockser) that "utilizing 7 bits of mantissa in multiplication operations resulting in a minimum of 12% of valid floating point 32 inputs producing at least 0.39% relative error compared to the exact mathematical calculation of a full-precision multiplication on those same inputs." PO Resp. 66–67 (citing Pet. 33–34, 72; Khatri Decl. ¶¶ 135–136). Patent Owner does not explain, however, how the cited analysis of how *Dockser* teaches the X and Y percentage limitations applies to the TPUv2 and TPUv3 products. As Petitioner points out, Patent Owner "does not even attempt to demonstrate that [Petitioner's] TPUs perform the same bit-dropping techniques as Dockser." *See* Reply 27. Patent Owner has not established a sufficient nexus between the alleged copying and commercial success of the TPUv2 and TPUv3 products and the Dockser/Tong-challenged claims.

Finally, we note that Patent Owner's evidence of commercial success would be unpersuasive even if Patent Owner had established a nexus. Patent Owner's sole support for the alleged commercial success of the TPUv2 and TPUv3 products is an article in *PCMag India*. *See* PO Resp. 62 (citing Ex. 2017). That article states in relevant part:

> As AI and Machine Learning advanced, [Petitioner] was faced with the proposition of doubling the number of data centers it operates to handle the increased workloads. Building multiple data centers to handle the growing amounts of data sets *could have easily cost the company anything above $10 billion*.

> So [Petitioner] decided to take another route by optimizing its current data centers for better efficiency. The search giant in late 2013 decided to produce a custom chip in order to improve

IPR2021-00155
Patent 10,416,961 B2

> cost-performance over the GPUs it had deployed. Thus *the first
> TPU* was born.

Ex. 2017, 3 (emphasis added). The above statement as to how much "could

have" been saved is too speculative to reach a conclusion of commercial

success. *See id.* It also pertains to Petitioner's "first" TPU, not the TPUv2

and TPUv3 products that use the bfloat16 format. *See id.*; Ex. 2016, 69

("Quantized arithmetic—8-bit integer instead of 32-bit floating point (FP)—

can work for inference like in TPUv1 but reduced-precision training is an

active research area.").

Accordingly, we find that Patent Owner's evidence of alleged copying

and commercial success does not weigh in favor of nonobviousness of the

Dockser/Tong-challenged claims.

g) Conclusion

Based on all the evidence of record, including evidence of objective

indicia of nonobviousness submitted by Patent Owner, we determine that

claim 1 would have been obvious based on Dockser under 35 U.S.C.

§ 103(a).

3. Claim 2

Claim 2 recites, "The device of claim 1, wherein the at least one first

computing device comprises at least one of a central processing unit (CPU),

a graphics processing unit (GPU), a field programmable gate array (FPGA),

a microcode-based processor, a hardware sequencer, and a state machine."

In Petitioner's view, only one of the alternatives recited in claim 2

needs to be met. Pet. 40. We agree because claim 2 recites "at least one of,"

which indicates that what follows should be read disjunctively.

51

IPR2021-00155
Patent 10,416,961 B2

Petitioner asserts that Dockser's "main processor" is a "general-purpose processor" implementable as a "state machine," as recited. *Id.* (citing Ex. 1007 ¶¶ 15, 35).

Patent Owner does not provide arguments specifically directed to this analysis. *See* PO Resp.; Sur-Reply.

We agree with Petitioner because Dockser's FPP can be part of the main processor, and the quoted part of Dockser explains that the FPP is part of a general-purpose processor. Ex. 1007 ¶¶ 15, 35. We credit the Goodin Declaration on these points, as it is explained sufficiently and consistent with the disclosures of the references. *See* Goodin Decl. ¶¶ 312–311. Based on all the evidence of record, including evidence of objective indicia of nonobviousness submitted by Patent Owner (*supra* Section II.D.2.f)), Petitioner has shown, by a preponderance of the evidence, that claim 2 would have been obvious based on Dockser under 35 U.S.C. § 103(a).

4. *Claim 10*

Claim 10 is an independent claim that recites, "A device comprising: a plurality of components comprising: at lease one first" LPHDR execution unit with all the same limitations as the LPHDR execution unit recited in claim 1. Ex. 1001, col. 31, ll. 25–43. That is, claim 10 is similar to claim 1 except that it adds "a plurality of components comprising" and lacks "at least one first computing device adapted to control the operation of the at least one first LPHDR execution unit" at the end. *Compare id.* (claim 10), *with id.* at col. 30, ll. 17–37 (claim 1).

According to Petitioner, Dockser's computing system has a plurality of components: a main processor and an FPP. Pet. 41 (citing Goodin Decl. ¶ 313). Petitioner, alternatively, argues that, if claim 10 were

IPR2021-00155
Patent 10,416,961 B2

interpreted to require that a plurality of components be included in a single LPHDR execution unit, Dockser's FPP has multiple components: registers, controller, and FPO, which also has adder and multiplier. *Id.* (citing Goodin Decl. ¶ 314; Ex. 1007 ¶¶ 15, 19–22). Petitioner relies on its analysis of claim 1 to address the remaining limitations, which are also found claim 1. *Id.* (citing Goodin Decl. ¶¶ 315–316).

Apart from its arguments directed to the features recited in claim 1, Patent Owner does not provide arguments specifically directed to this analysis. *See* PO Resp.; Sur-Reply.

We agree with Petitioner that Dockser's computing system has a plurality of components under either interpretation. *See* Pet. 41. Those components are described in the cited parts of Dockser. Ex. 1007 ¶¶ 15, 19–22. We credit the Goodin Declaration on these points, as it is explained sufficiently and consistent with the disclosures of the references. *See* Goodin Decl. ¶¶ 313–314. For the reasons discussed in Section II.D.2, Petitioner has shown that Dockser teaches the features recited in claim 10 that are shared with claim 1.

Based on all the evidence of record, including evidence of objective indicia of nonobviousness submitted by Patent Owner (*supra* Section II.D.2.f)), Petitioner has shown, by a preponderance of the evidence, that claim 10 would have been obvious based on Dockser under 35 U.S.C. § 103(a).

5. *Claims 4 and 13*

Claim 4 recites, "The device of claim 1, wherein the dynamic range of the possible valid inputs to the first operation is at least as wide as from 1/1,000,000 through 1,000,000." Ex. 1001, col. 30, ll. 49–51. Claim 13

IPR2021-00155
Patent 10,416,961 B2

recites a similar limitation but depends from claim 10. *Id.* at col. 31, ll. 52–54.

According to the Petition, "[t]he dynamic range of possible valid inputs to Dockser's operation extends from approximately 2^{-126} (much smaller than 1/1,000,000) to approximately 2^{127} (much larger than 1,000,000), meeting claims 4 and 13." Pet. 41 (citing Goodin Decl. ¶¶ 317–318; Pet., Section V.B.3 (discussing claim 1)).

Patent Owner does not provide arguments specifically directed to this analysis. *See* PO Resp.; Sur-Reply.

We agree with Petitioner that Dockser's dynamic range meets the recited range. Pet. 41. The dynamic range is determined by the number of exponent bits. *See, e.g.*, Ex. 1015, col. 8, ll. 48–49 ("The dynamic range increases with the number of bits dedicated to the exponent . . . ."), *cited in* Goodin Decl. ¶ 231, Pet. 20. Dockser's FPP operates on IEE-745 numbers with 8-bit exponents. Ex. 1007 ¶¶ 2, 16, 17, 24, *cited in* Pet. 20 (Section V.B.3). Those 8-bit exponents range from -126 to 127, which means that the dynamic range of "normal" IEEE-754 single-format operands is approximately 2^{-126} to 2^{127}." Goodin Decl. ¶ 232.

We credit the Goodin Declaration on these points, as it is explained sufficiently and consistent with the disclosures of the references. *See* Goodin Decl. ¶¶ 232–231, 317–318.

Based on all the evidence of record, including evidence of objective indicia of nonobviousness submitted by Patent Owner (*supra* Section II.D.2.f)), Petitioner has shown, by a preponderance of the evidence, that claims 4 and 13 would have been obvious based on Dockser under 35 U.S.C. § 103(a).

IPR2021-00155
Patent 10,416,961 B2

6.   Claims 5 and 14

Claim 5 recites, "The device of claim 1, wherein the first operation is multiplication." Claim 14 recites a similar limitation but depends from claim 10. Petitioner asserts that "Dockser's claimed 'first operation' is reduced-precision 'multiplication.'" Pet. 41 (citing Goodin Decl. ¶¶ 319–321). Patent Owner does not provide arguments specifically directed to this analysis. *See* PO Resp.; Sur-Reply. We agree with Petitioner for the reasons discussed in Section II.D.2, where we analyzed the "first operation" in the context of claim 1.

We credit the Goodin Declaration on these points, as it is explained sufficiently and consistent with the disclosures of the references. Goodin Decl. ¶¶ 319–321.

Based on all the evidence of record, including evidence of objective indicia of nonobviousness submitted by Patent Owner (*supra* Section II.D.2.f)), Petitioner has shown, by a preponderance of the evidence, that claims 5 and 14 would have been obvious based on Dockser under 35 U.S.C. § 103(a).
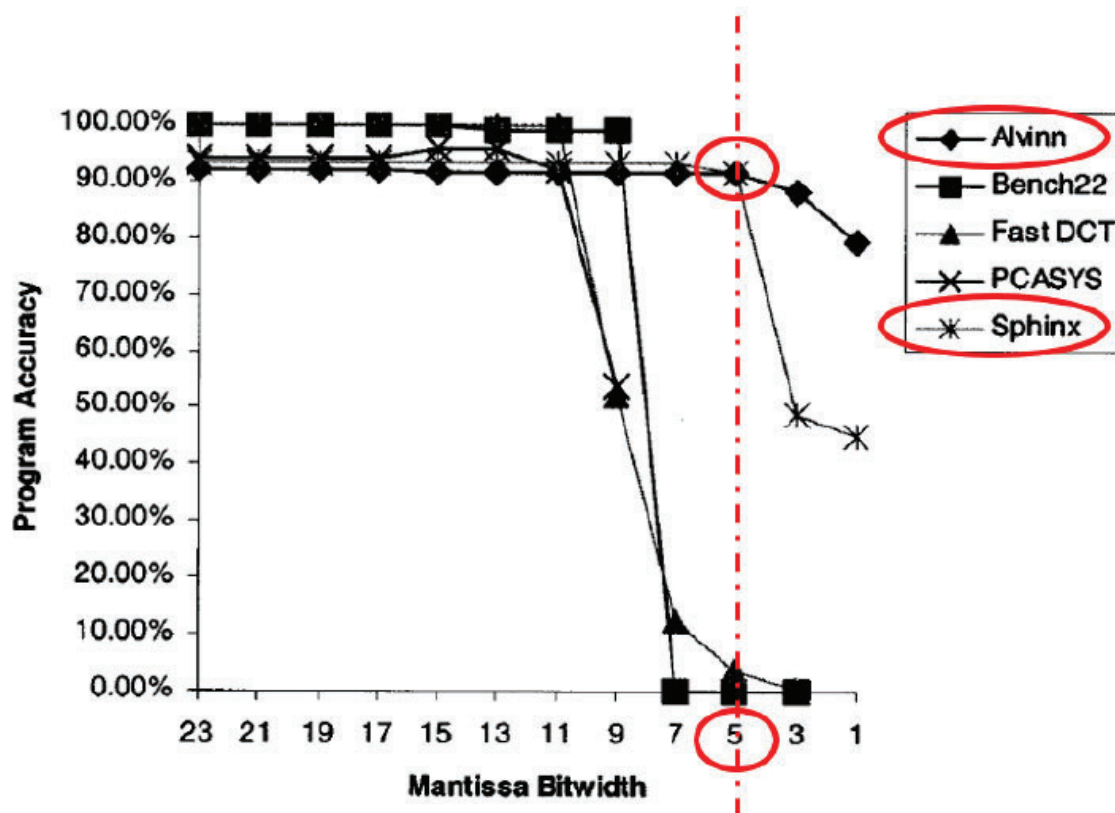
E.   Obviousness Ground Based on Dockser and Tong (Claims 1, 2, 4, 5, 10, 13, 14, 21, 24, and 25)

1.   Tong

Tong is an IEEE journal article entitled "Reducing Power by Optimizing the Necessary Precision/Range of Floating-Point Arithmetic." Ex. 1008, 273. Tong teaches reducing power consumption by minimizing the bitwidth representation of floating-point data. *Id.* According to Tong, using a variable bitwidth floating-point unit saves power. *Id.*

55

IPR2021-00155
Patent 10,416,961 B2

2. *Claim 1*

Petitioner argues that the subject matter of claim 1 would have been obvious over the combination of Dockser and Tong for reasons similar to those analyzed with respect to Dockser alone. *See* Pet. 42–45. Additionally, Petitioner asserts that "Tong, like Dockser . . . , confirms that the number of mantissa bits used in a high-dynamic-range floating-point execution unit was a well-known result-effective variable impacting power consumption and precision." *Id.* at 43 (citing Ex. 1008, 273–278, Goodin Decl. ¶¶ 323–325). Petitioner relies on Tong's Figure 6, reproduced below with Petitioner's annotations. *Id.* at 44.

Tong's Figure 6, above, is a line graph showing program accuracy, from 0% to 100%, on the vertical axis and mantissa bitwidth, from 1 to 23, on the horizontal axis. *Id.* at 43–44. The figure shows this data for five programs:

IPR2021-00155
Patent 10,416,961 B2

ALVINN, Bench22, Fast DCT, PCASYS, and Sphinx III. *Id.* The programs implement different signal-processing tasks. *See, e.g.*, Ex. 1008, 278 (Table IV). ALVINN, for example, is a neural network trainer that uses backpropagation. *Id.* And Sphinx III is a speech-recognition program. *Id.*

Petitioner argues that "Tong's teaching that a precision level retaining 5 mantissa fraction bits is sufficient in some applications (including ALVINN and Sphinx [III]) would have motivated a [person of ordinary skill in the art] to configure Dockser's FPP . . . to operate at a selected precision level retaining as few as 5 mantissa fraction bits." *Id.* at 45. According to the Petition, one of ordinary skill in the art would have done so "to conserve power when running those applications, or others empirically determined (using Tong's techniques) to not require greater precision." *Id.* (citing Goodin Decl. ¶ 333). Petitioner also concludes that determining the "optimum range of imprecision to achieve the best power reduction without sacrificing accuracy for a particular application" was a matter of routine optimization of a result-effective variable. *Id.* at 46–47.

Patent Owner argues that, because Dockser lacks an LPHDR execution unit and Tong does not remedy that deficiency, the Dockser-Tong combination does not render obvious any challenged claim. PO Resp. 24; Sur-Reply 9. We disagree for the reasons discussed in Section II.D.2 analyzing Petitioner's challenge based on Dockser alone.

Patent Owner further argues that Tong does not teach using only 5 mantissa bits, which means that the Dockser-Tong combination lacks the imprecision required to meet the X and Y values recited in the imprecision limitation. PO Resp. 24. In Patent Owner's view, "Tong teaches that at least 11 mantissa bits are required for consistent performance, even for 'programs dealing with human interfaces [that] process sensory data with

IPR2021-00155
Patent 10,416,961 B2

intrinsically low resolutions.'" *Id.* at 25 (citing Ex. 1008, 278). According

to Patent Owner, "Petitioner makes no effort to show that arithmetic with the

'required' 11-bit mantissa is an LPHDR operation and meets the X/Y

limitations of the claims." PO Resp. 25.

Yet Patent Owner's argument about an 11-bit mantissa (PO Resp. 24–

25) does not squarely address Petitioner's obviousness rationale—i.e., that

one of ordinary skill in the art would have been motivated to configure

Dockser's FPP to operate at a "selected precision level" according to Tong

(Pet. 45). Specifically, Petitioner's rationale proposes using a selected

precision level "retaining as few as 5 mantissa fraction bits." Pet. 45. Thus,

we disagree with Patent Owner's argument about an 11-bit mantissa. PO

Resp. 24–25; Sur-Reply 9. For this reason, we do not credit the Khatri

Declaration on this point. Khatri Decl. ¶ 77.

Petitioner used a software program to determine the relative error

when retaining 5 mantissa fraction bits—i.e., the number of bits that

Dockser would use under Petitioner's proposed combination with Tong.

Pet. 46 (citing Goodin Decl. ¶ 335). Also, the algebraic analysis cited by

Petitioner uses 5 mantissa fraction bits. *Id.* (citing Goodin Decl. ¶ 336).
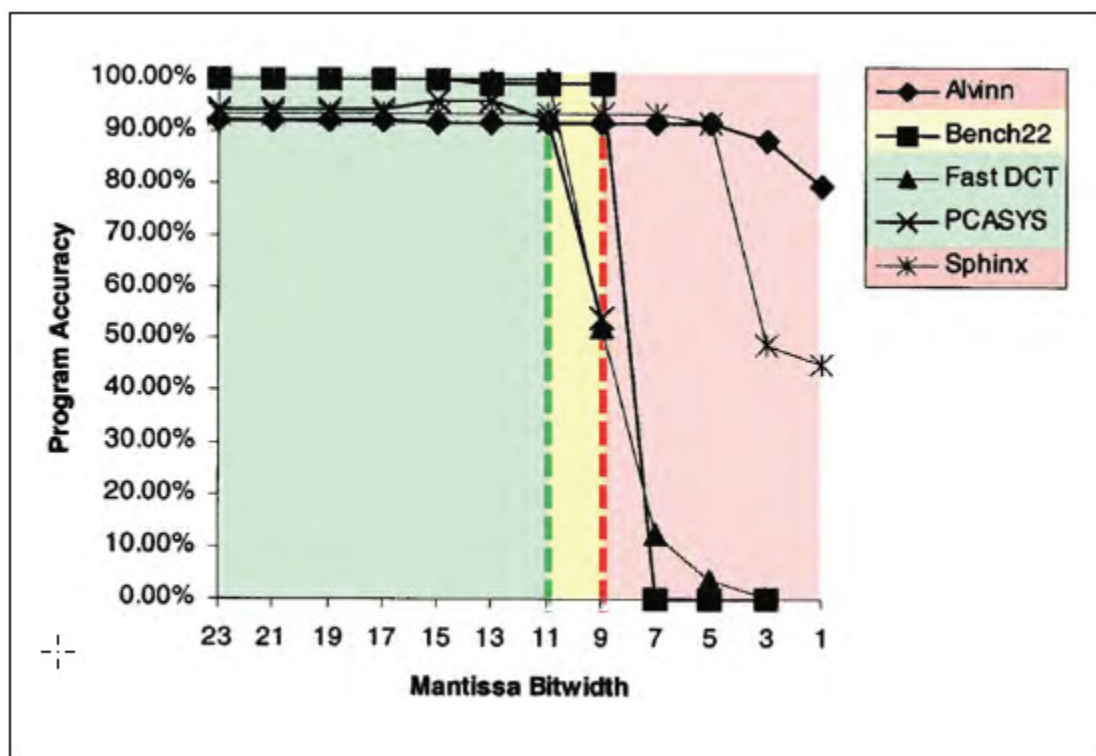
Petitioner adequately shows that the X and Y limitations are met for

the 5-mantissa bit rationale. Pet. 46. In particular, Mr. Goodin's software

program shows "when retaining 5 mantissa fraction bits in view of Tong,

Dockser's register bit-dropping technique produces at least Y=0.2% relative

error for 99.03% of possible valid inputs, and Dockser's logic bit-dropping

technique produces at least Y=0.2% relative error for 99.44% of possible

valid inputs," and the algebraic analysis shows that "Dockser's register bit-

dropping technique when retaining 5 mantissa fraction bits produces a

minimum of 1.56% relative error (greater than Y=0.2%) for over 12% of

IPR2021-00155
Patent 10,416,961 B2

possible valid inputs (greater than X=10%)." Goodin Decl. ¶ 335. After

review of the arguments and evidence on these issues, we credit the Goodin

Declaration's algebraic analysis and software-based analysis. *Id.* ¶ 335,

Appendix I.B–D.

  Although Patent Owner does not present argument specifically

directed to Mr. Goodin's software-based or algebraic analysis, Patent Owner

argues that one of ordinary skill in the art would not have chosen a 5-bit

mantissa: "Tong's experimental results show that using fewer than 11

mantissa bits unacceptably reduces accuracy for the benchmark applications

that were specifically selected (from a category of programs) for their high

tolerance for imprecision." PO Resp. 25 (citing Khatri Decl. ¶ 78). To

illustrate this point, Patent Owner annotates Tong's Figure 6, reproduced

below:

Tong's Figure 6, above, shows the experimental results with color

annotations indicating "the precision levels suitable for general-purpose

IPR2021-00155
Patent 10,416,961 B2

operation (green), those at which a significant percentage of applications begin to produce unacceptable results (yellow), and finally, levels of precision that are unsuitable for most testing benchmarks (red)." *Id.* at 26 (citing Khatri Decl. ¶ 79).

Patent Owner argues that there is no motivation to combine Dockser and Tong because "there is no evidence to show that a POSA [person of ordinary skill in the art] would have chosen Tong's 5-bit examples over the 7, 9, or 11-bit examples." *Id.* at 26–27; *see also* Sur-Reply 10. According to Patent Owner, "Tong does not disclose any extra capability that would have motivated a POSA to use it with Dockser." PO Resp. 27 (citing Khatri Decl. ¶¶ 81–82). In Patent Owner's view, Petitioner relies on hindsight to choose 5 bits and does not offer any differences between Dockser and the claimed invention or other non-hindsight motivation to combine with Tong. *Id.* According to Patent Owner, "Tong teaches that 'it appears inevitable that some fraction of our operands will require full IEEE standard precision,'" which means that a person of ordinary skill in the art would not have been motivated to operate only in 5-bit, 7-bit, or 9-bit modes as Petitioner suggests. Sur-Reply 10.

We disagree with Patent Owner's arguments directed to the rationale for using 5 mantissa bits. *See* PO Resp. 25–27; Sur-Reply 10. Petitioner's reasoning for the 5-bit example is "to conserve power when running [certain] applications, or others empirically determined (using Tong's techniques) to not require greater precision." Pet. 45 (citing Goodin Decl. ¶ 333). For example, in the ALVINN and Sphinx III line plots in Tong's Figure 6, Petitioner added a dashed red line extending vertically through the data points with a mantissa bitwidth of 5. *Id.* at 44. Petitioner asserts that, for these programs, Tong teaches that "the accuracy does not

60

IPR2021-00155
Patent 10,416,961 B2

change significantly with as few as 5 mantissa [fraction] bits." *Id.* (quoting Ex. 1008, 278, § V.B.). In Petitioner's view, Tong omits unnecessary bits to reduce waste and power consumption. *Id.* Petitioner characterizes Tong as "[h]aving empirically determined the minimum number of mantissa bits necessary to maintain acceptable accuracy of particular applications." *Id.* (citing Ex. 1008, 273, 274, 279, 284).

Petitioner's assertions are adequately supported by the record. We are persuaded by Petitioner's reliance on Tong's Figure 6 to determine how many bits are needed in a particular application. *See* Pet. 44. That is, Petitioner's choice is not arbitrary or solely based on the '961 patent's written description alone. *See* PO Resp. 25–27. Rather, Petitioner's choice of 5-bits is based on Tong's experimental results for *certain* applications. *See* Pet. 44. We emphasize *certain* because Petitioner is not saying that 5-bits is suitable for every application. *See id.* Full precision may be needed some of the time. *See* Ex. 1008, 280, *discussed in* Sur-Reply 10. But, even if Tong teaches that some applications begin to produce unacceptable results below 11-bits, as Patent Owner notes (PO Resp. 25–26), Tong teaches that 5-bits would be enough for certain ones: ALVINN and Sphinx III shown in Tong's Figure 6. Pet. 44.

Patent Owner argues that "Tong lists those two applications amongst a great many other disclosed applications that all require full precision, and that a POSA would therefore not be motivated to configure Dockser to operate with a 5-bit mantissa just to support those two applications given the far larger number of disclosed applications that require full precision." Sur-Reply 9; PO Resp. 24-27.

We disagree with Patent Owner's argument here because the record adequately supports Petitioner's rationale that "[d]etermining the optimum

61

IPR2021-00155
Patent 10,416,961 B2

range of imprecision to achieve the best power reduction without sacrificing accuracy for a particular application" was a matter of routine optimization of a result-effective variable, and Tong, in particular, shows an optimization for a particular program. Pet. 46–47. The record adequately supports Petitioner's reasoning here: Dr. Khatri testified that "exact computing . . . takes up a lot of power." Ex. 1072 ("Deposition of Sunil P. Khatri"), 47:2–5. Mr. Goodin explained that reducing precision may reduce power consumption. *See, e.g.*, Ex. 2043 ("Deposition of Richard M. Goodin"), 34:21–35:4; *see also id.* at 31:2–32:1 (discussing how selectable precision may achieve savings). And we credit the Goodin Declaration's explanation of how Tong achieves this for certain applications. *See, e.g.*, Goodin Decl. ¶¶ 31–33 (citing Ex. 1008, 273, 277–279), 333, 336 (citing Ex. 1008, 278). That is, Petitioner's proposed combination would omit unnecessary bits to reduce waste and power consumption. Pet. 44.

Thus, we disagree with Patent Owner's arguments regarding the combination with Tong. *See* PO Resp. 25–27; Sur-Reply 9–10. For the same reasons, we do not credit the Khatri Declaration on this point. Khatri Decl. ¶¶ 74–82.

For the reasons discussed, we determine that Petitioner provides articulated reasoning, supported by rational underpinnings, why one of ordinary skill in the art would have combined Dockser and Tong. *See KSR*, 550 U.S. at 418. Based on the Petition, we find that a person of ordinary skill in the art would have modified Dockser with Tong to arrive at the limitations recited in claim 1.

Thus, Petitioner has shown, by a preponderance of the evidence, that claim 1 would have been obvious based on Dockser and Tong under 35 U.S.C. § 103(a) with a reasonable expectation of success. In coming to this

IPR2021-00155
Patent 10,416,961 B2

conclusion, we also have taken into account the parties' arguments and evidence regarding the objective indicia of nonobviousness. *Supra* Section D.2.f).

3. *Claims 2, 4, 5, 10, 13, and 14*

Petitioner asserts that the subject matter recited in claims 2, 4, 5, 10, 13, and 14 would have been obvious over the combination of Dockser and Tong for reasons similar to those presented in the ground based on Dockser alone. *See* Pet. 45–47. For the reasons discussed in Section II.D.2 analyzing Dockser alone, Petitioner's contentions are supported by the Goodin Declaration and are persuasive. *See id.* Patent Owner does not present arguments specifically directed to the challenge based on the Dockser-Tong challenge to these claims. *See* PO Resp.; Sur-Reply.

To the extent Patent Owner's arguments apply to claims 2, 4, 5, 10, 13, and 14, we disagree with Patent Owner's arguments for the same reasons explained above. *See supra* Sections II.D.2, II.E.2. After reviewing the arguments and evidence on these issues, we agree with Petitioner and credit Mr. Goodin's testimony. Goodin Decl. ¶¶ 322–346. We determine that Petitioner provides articulated reasoning, supported by rational underpinnings, why one of ordinary skill in the art would have combined Dockser and Tong. *See KSR*, 550 U.S. at 418. Based on the Petition, we find that a person of ordinary skill in the art would have modified Dockser with Tong to arrive at the subject matter recited in claims 2, 4, 5, 10, 13, and 14 with a reasonable expectation of success. Based on all of the evidence of record, including evidence of secondary considerations of nonobviousness submitted by Patent Owner, we determine that the subject matter of claims

IPR2021-00155
Patent 10,416,961 B2

2, 4, 5, 10, 13, and 14 would have been obvious based on Dockser and Tong

under 35 U.S.C. § 103(a). *Supra* Section D.2.f).

4. *Claim 21*

Claim 21 recites limitations similar to those in claim 1 but as a

computer program to emulate a device:

> A device comprising a computer processor and a computer-readable memory storing computer program instructions, wherein the computer program instructions are executable by the processor to emulate a second device comprising: a plurality of components comprising: at least one first low precision high-dynamic range (LPHDR) execution unit adapted to . . . .

Ex. 1001, col. 32, ll. 37–44.

Petitioner asserts that Tong teaches emulating, in software, different

bitwidth FP units to determine application accuracy. Pet. 47 (citing

Ex. 1008, 278). According to Petitioner, based on this teaching, one of

ordinary skill in the art would have been motivated "to emulate the

Dockser/Tong device ([Pet.] § VI.B) in software to assess the accuracy of

applications running on the device at selected precision levels, and a POSA

would have had a reasonable expectation of success in doing so." *Id.* (citing

Goodin Decl. ¶¶ 339–340).

Petitioner asserts that "[w]hen the Dockser/Tong device is 'emulated

in software' per Tong's teachings ([Ex. 1008], 278) cited [in Pet.] § VI.A,

the conventional and obvious implementation of such 'software' is

'computer program instructions' stored in 'a computer-readable memory'

and 'executable by [a] processor' within a 'device' (e.g., a computer) as

claimed." *Id.* at 48 (citing Goodin Decl. ¶ 341; Ex. 1007 ¶ 36;

Ex. 1029 ¶¶ 8, 9, 14; Ex. 1001, col. 26, ll. 13–15).

IPR2021-00155
Patent 10,416,961 B2

For reasons similar to those with respect to claim 1, Petitioner argues that a person of ordinary skill in the art would have been motivated by Tong to use a computer to emulate the combined teachings of Dockser and Tong, "meeting the 'second device' of [21A1]), which comprises a plurality of components including a main processor and Dockser's execution unit (which itself comprises multiple components)." *Id.* (citing Goodin Decl. ¶ 342). Petitioner argues that Tong emulates a device in the same way as the '961 patent does:

> Like the '961 patent's only examples of "emulat[ing] a… device" comprising LPHDR execution unit(s), Tong emulates the device by executing a software application program that runs thereon with arithmetic operations in the application program replaced by software simulations of the reduced-precision arithmetic the device would perform.

Id. (citing Ex. 1001, col. 17, ll. 14–28, col. 18, l. 65–col. 19, l. 25; Ex. 1008, 278; Goodin Decl. ¶¶ 343).

Patent Owner does not provide arguments specifically directed to Petitioner's arguments or evidence about claim 21 or software emulation. *See* PO Resp.; Sur-Reply.

After review of the arguments and evidence on these issues, we agree with Petitioner and credit the Goodin Declaration. Goodin Decl. ¶¶ 341, 343.

In particular, the '961 patent describes code written in the C programming language to perform the same arithmetic operations in the same order using the same method as one implemented in hardware. Ex. 1001, col. 18, l. 65–col. 19, l. 25, *cited in* Pet. 48. We agree with Petitioner that Tong similarly uses a computer program to simulate reduced-precision arithmetic that a hardware unit would perform. Pet. 48. For

65

IPR2021-00155
Patent 10,416,961 B2

example, Tong uses an "FP software emulation package" to simulate reduced-precision arithmetic. Ex. 1008, 278, *cited in* Pet. 48.

Petitioner's obviousness rationale—i.e., "to assess the accuracy of applications running on the device at selected precision levels"—is adequately supported by the record: Tong explains that the emulation package was used to assess "different mantissa and exponent bitwidths." *See* Pet. 47 (citing Ex. 1008, 278). Petitioner's assertion that software is computer program instructions stored in a computer-readable memory and executable by a processor within a computing device is adequately supported by at least the Goodin Declaration and the evidence that it cites. Pet. 48 (citing Goodin Decl. ¶ 341; Ex. 1007 ¶ 36; Ex. 1029 ¶¶ 8, 9, 14; Ex. 1001, col. 26, ll. 13–15).

For the same reasons discussed with respect to claim 1, Petitioner shows the subject matter recited in claim 21 that is also recited in claim 1 would have been obvious over the Dockser-Tong combination. *See supra* Section II.E.2.

Thus, we determine that Petitioner provides articulated reasoning, supported by rational underpinnings, why one of ordinary skill in the art would have combined Dockser and Tong. *See KSR*, 550 U.S. at 418. Based on the Petition, and taking into account Patent Owner's arguments on objective indicia, we find that a person of ordinary skill in the art would have modified Dockser with Tong to arrive at the limitations recited in claim 21.

Accordingly, Petitioner has shown that claim 21 is unpatentable over the Dockser-Tong combination.

IPR2021-00155
Patent 10,416,961 B2

5.   *Claims 24 and 25*

Claims 24 and 25 depend from claim 21 but are otherwise identical to claims 4 and 5. Pet. 49. Patent Owner does not present arguments specifically directed to these claims. *See* PO Resp.; Sur-Reply. For the same reasons discussed with respect to claims 4, 5, and 21, Petitioner has shown that claims 24, and 25 are unpatentable over the Dockser-Tong combination, and we credit the corresponding parts of the Goodin Declaration. Pet. 47–49; *see* Goodin Decl. ¶¶ 339–346.

F.   *Obviousness Ground Based on Dockser and MacMillan (Claim 3)*

1.   *MacMillan*

MacMillan is entitled "Circuit for Enhancing Performance of a Computer for Personal Use." Ex. 1009, code (54). MacMillan discloses using Single Instruction Multiple Data (SIMD) parallel-processing architectures for "adding supercomputer performance to a computer for personal use." *Id.* at col. 5, ll. 22–54. MacMillan's computer system comprises a "Host CPU" (i.e., "a 386, 486 or Pentium[] processor") and SIMD-random access memory (SIMD-RAM) devices. *Id.* at col. 9, ll. 11–19, 30–31, Figs. 2, 3. Figure 5 of MacMillan is reproduced below.

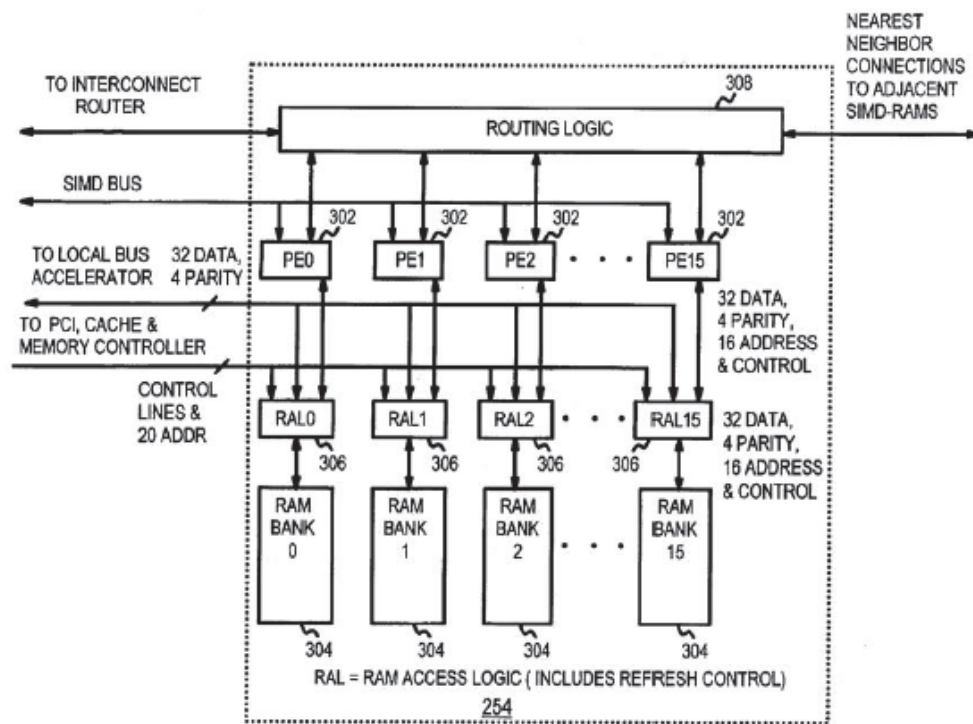IPR2021-00155
Patent 10,416,961 B2

FIG. 5

Figure 5 depicts SIMD-RAM device 254 comprising a plurality of
DRAMs 304 coupled to a plurality of processing elements 302 (PE0, PE1,
PE2, etc.) connected to SIMD bus 240. *Id.* at col. 12, ll. 35–43. MacMillan
discloses that "[e]ach PE contains a 32-bit wide data path and can perform
atomic operations on bits, bytes, 16-bit words, and 32-bit words," "each
DRAM 304 is independently addressed by its PE," "[i]nteger and floating
point accelerators could be included in each PE," and "[e]xecution autonomy
is provided, in which specific PEs can be excluded from executing specific
instructions." *Id.* at col. 12, ll. 47–59. MacMillan describes an example
architecture of a system with 256 PEs, but states that the disclosed
architecture "allows scaling to higher or lower density chips with more or
fewer PEs." *Id.* at col. 12, l. 60–col. 13, l. 4, col. 13, ll. 38–41, col. 16,
ll. 20–22.

IPR2021-00155
Patent 10,416,961 B2

2. Claim 3

Petitioner contends that claims 1–5, 10, 13, and 14 are unpatentable over Dockser and MacMillan under 35 U.S.C. § 103(a). Pet. 49–55. Of those claims, claims 1 and 10 are independent. Petitioner has shown that claims 1, 2, 4, 5, 10, 13, and 14 are unpatentable. *See supra* Section II.D. Given our disposition of the grounds based on Dockser and the combination of Dockser and Tong, we do not reach Petitioner's alternative ground asserting that claims 1, 2, 4, 5, 10, 13, and 14 are unpatentable over Dockser and MacMillan. *See Boston Sci. Scimed, Inc. v. Cook Grp. Inc.*, 809 F. App'x 984, 990 (Fed. Cir. Apr. 30, 2020) (non-precedential) (recognizing that the "Board need not address issues that are not necessary to the resolution of the proceeding" and, thus, agreeing that the Board has "discretion to decline to decide additional instituted grounds once the petitioner has prevailed on all its challenged claims"). Claim 3 is only challenged under this ground and the ground based on Dockser, MacMillan, and Tong. *See infra* Section II.G.

We are not persuaded that a person of ordinary skill in the art would have been motivated to combine the teachings of Dockser and MacMillan in the manner asserted by Petitioner. Thus, Petitioner has not shown, by a preponderance of the evidence, that claim 3 would have been obvious based on Dockser and MacMillan under 35 U.S.C. § 103(a). Our reasoning follows.

Claim 3 depends from claim 1 and recites that "the number of LPHDR execution units in the device exceeds by at least one hundred the non-negative integer number of execution units in the device adapted to execute at least the operation of multiplication on floating point numbers that are at least 32 bits wide" (the "exceeds" limitation). Petitioner relies on

IPR2021-00155
Patent 10,416,961 B2

MacMillan for its teachings regarding multiple floating-point execution units, and argues that a person of ordinary skill in the art "would have been motivated to use Dockser's FPP to implement each 'floating-point accelerator' in the parallel PEs of MacMillan's SIMD architecture to increase performance speed as MacMillan teaches while lowering power consumption as Dockser teaches." Pet. 51–52 (citing Goodin Decl. ¶¶ 353–354). According to Petitioner, "using Dockser's FPP as the floating-point accelerator in MacMillan's PE would have achieved the predictable result of enabling the PEs to perform reduced-precision floating-point arithmetic as taught by Dockser at reduced power." *Id.* at 52 (citing Goodin Decl. ¶ 354; Ex. 1009, col. 12, ll. 47–49; Ex. 1024, col. 1, ll. 43–58). Petitioner further contends that MacMillan's SIMD parallel processing is used in "the same types of computers for which Dockser's FPP provides '[p]ower management'" and would be beneficial for the same types of "'graphics applications' for which Dockser teaches its FPP can beneficially save power by reducing unnecessary precision." *Id.* at 51–52 (citing Goodin Decl. ¶¶ 351, 355; Ex. 1007 ¶ 3; Ex. 1009, col. 1, ll. 6–9, col. 5, ll. 22–45, col. 7, ll. 14–34).

With respect to the "exceeds" limitation in particular, Petitioner asserts that MacMillan teaches 256 PEs, and the Dockser-MacMillan combination would have a single Host CPU and at least one FPP in each PE. *Id.* at 54 (citing Goodin Decl. ¶¶ 372–373). Petitioner asserts that MacMillan's example architecture includes 256 PEs and scales to tens of thousands or more. *Id.* (citing Ex. 1009, col. 12, l. 60–col. 13, l. 4; col. 13, ll. 39–41; col. 16, ll. 20–22). Thus, in Petitioner's proposed combined device based on Dockser and MacMillan, the number of LPHDR execution units (Dockser FPPs) "exceeds by over 100 the non-negative integer number

IPR2021-00155
Patent 10,416,961 B2

of Host CPU floating-point execution units." *Id.* (citing Goodin Decl.
¶¶ 372–373).

Patent Owner responds that a person of ordinary skill in the art would
not have been motivated to combine the teachings of Dockser and
MacMillan in the manner asserted by Petitioner. PO Resp. 35–36. Patent
Owner argues that "Dockser is focused on the objective of reducing power
consumption," whereas "MacMillan is focused on a parallel architecture that
increases computational power, and not focused on reducing power
consumption." *Id.* at 35 (citing Ex. 1009, col. 3, ll. 4–6; Khatri Decl. ¶ 102).
Patent Owner further contends that "[i]ncorporating Dockser's FPPs into
MacMillan would . . . defeat MacMillan's stated objective of achieving a
high-scale SIMD computer architecture at 'lower system cost.'" *Id.* at 36
(citing Ex. 1009, col. 5, ll. 58–59; Khatri Decl. ¶ 103). According to Patent
Owner, "Dockser's FPPs are even *larger* than traditional full-precision
execution units because of the control circuitry needed to implement the
selectable subprecision modes," such that "replacing the full-precision
execution units of MacMillan with Dockser FPP units would require
additional circuitry and chip space and would therefore increase costs, while
providing no benefit." *Id.* (citing Khatri Decl. ¶ 104); *see also* Sur-Reply
14–15.

Petitioner argues in its Reply that the asserted combination would
achieve both benefits of increasing performance speed (based on
MacMillan) and lowering power consumption (based on Dockser), and that
Patent Owner fails to provide supporting evidence for its argument that
incorporating Dockser's FPP into the MacMillan device would increase
costs, and even if it had, economic reasons are not indicative of
nonobviousness. Reply 9–10 (citing Ex. 1071 ¶¶ 30–31).

71

IPR2021-00155
Patent 10,416,961 B2

After reviewing the parties' arguments and evidence, and the full disclosures of Dockser and MacMillan, we agree with Patent Owner. Dockser teaches a single FPP with "selectable subprecision" that is capable of both full precision and reduced precision operations. *See* Ex. 1007, code (57), ¶ 15. The FPP includes additional control circuitry, such as floating-point controller 130 and control register 137, that allows selection of a subprecision for floating-point operations. *Id.* ¶¶ 15, 18, Fig. 1. Dockser describes "a variety of ways" in which "subprecision select bits may be used to reduce the precision of a floating-point operation." *Id.* ¶¶ 18, 23–34. Reduced precision can be achieved, with resultant power savings, either by removing power from "floating-point register elements for the excess bits of the fraction that are not required to meet the precision specified by the subprecision select bits" or "removing power to the logic in the floating-point operator 140 that remains unused as a result of the subprecision selected." *Id.* ¶¶ 4–7, 26–27. Mr. Goodin agrees that Dockser's FPP requires additional control circuitry to implement its selectable subprecision. Ex. 2043, 43:21–45:9.

MacMillan, by contrast, is heavily focused on avoiding increased costs associated with added complexity. MacMillan states that "[i]f supercomputing performance could be achieved in low cost computers, such as personal computers, this could dramatically expand the market for personal computers." Ex. 1009, col. 1, ll. 21–24. SIMD computing capability can be added to personal computers, "resulting in much higher performance computing at moderate cost." *Id.* at col. 8, ll. 51–55. MacMillan, however, describes numerous limitations of known SIMD designs at the time, many of which relate to increased costs from added complexity (e.g., additional registers), size, etc. *See, e.g., id.* at col. 2,

IPR2021-00155
Patent 10,416,961 B2

ll. 56–61 ("hav[ing] a large register set on each PE . . . increases die area per

PE, resulting in higher costs per PE"), col. 2, ll. 62–65 ("Adding pins can

reduce the PE-to-memory bottleneck, but leads to increased packaging costs.

It may also require tighter geometries or increased numbers of layers on

printed circuits boards, further increasing costs."), col. 3, ll. 1–3 ("Adding

output buffers to drive increased pin counts also increases power dissipation

and hence power supply capacity and cost."), col. 4, ll. 39–40 ("The PE

array[] disk may be limited in size due to technology or cost factors.").

> Importantly, MacMillan states:

> Overcoming the above limitations would allow supercomputing
> performance to be provided in a *low cost* computer system for
> personal use, dramatically expanding the potential market for
> systems with supercomputer performance.   To meet the cost
> objectives, the SIMD capabilities *should not add significant
> complexity to the architecture of a computer system for personal
> use*.   The present invention addresses the above needs.

Id. at col. 5, ll. 38–45 (emphasis added).   MacMillan's disclosed

architecture—with multiple, less costly PEs operating in parallel to increase

computational power—is expressly intended to account for these limitations.

See id. at col. 5, ll. 44–45, col. 5, ll. 58–61 (the disclosed system with

"shared memory results in lower system cost and more flexibility in using

memory"), col. 6, ll. 1–10 ("Another aspect of the invention is that it shows

how parallel SIMD processing can be added inexpensively to existing

system architectures.   This cost-effective addition of parallel processing can

be made to popular, low cost computer systems running popular operating

systems . . . ."), col. 6, ll. 41–45, col. 16, ll. 14–27.

In determining how a person of ordinary skill in the art would have

understood those disclosures, we credit the testimony of Dr. Khatri, which is

consistent with, and a logical reading of, Dockser and MacMillan.   *See Icon*

IPR2021-00155
Patent 10,416,961 B2

Health & Fitness, Inc. v. Strava, Inc., 849 F.3d 1034, 1041 (Fed. Cir. 2017)
(stating that the Board "is permitted to weigh expert testimony and other
record evidence and, in so doing, rely on certain portions of an expert's
declaration while disregarding others"). Specifically, the circuitry
implementing Dockser's selectable subprecision is beyond that of a typical
full precision execution unit. Dr. Khatri testifies that

> [b]ecause Dockser's execution unit is capable of
> full-precision, and uses conventional arithmetic units, a [person
> of ordinary skill in the art] would understand that it would have
> at least as many transistors and take up at least as much space as
> a conventional full-precision arithmetic unit, even when
> operating in a reduced-precision mode. Indeed, because
> Dockser's execution unit includes additional control circuits for
> selecting reduced-precision modes, a [person of ordinary skill in
> the art] would expect it to be *larger* than a conventional
> full-precision execution unit, making it unsuitable for scaling,
> and therefore unsuitable for use in parallel processing arrays.

See Khatri Decl. ¶ 51. Petitioner's asserted combination of incorporating
Dockser's FPP into each of multiple PEs in MacMillan's device "would
require additional circuitry and chip space and would therefore *increase*
costs," "defeat[ing] MacMillan's stated objective of achieving a highly
parallel SIMD computer architecture at 'lower system cost.'" *See id.*
¶¶ 103–104. Again, MacMillan specifically counsels that to achieve its cost
objectives, "SIMD capabilities should not add significant complexity to the
architecture of a computer system for personal use." Ex. 1009, col. 5,
ll. 42–44. The combination proposed by Petitioner would do so.
MacMillan's PEs already perform full precision operations; Dockser's
selectable subprecision circuitry would add complexity beyond that of a
typical full precision execution unit. *See* Khatri Decl. ¶¶ 103–104; Ex. 1009,

IPR2021-00155
Patent 10,416,961 B2

col. 12, ll. 47–49 ("Each PE contains a 32-bit wide data path and can perform atomic operations on . . . 32-bit words.").

A person of ordinary skill in the art, reading MacMillan and understanding its emphasis on avoiding increased costs associated with added complexity and size, would not have been motivated to modify MacMillan's device by incorporating Dockser's FPP into the PEs. *See Chemours Co. FC, LLC v. Daikin Indus., Ltd.*, 4 F.4th 1370, 1376–77 (Fed. Cir. 2021) (finding insufficient motivation to "increase [a prior art reference's] melt flow rate to the claimed range," in part because the reference "includes numerous examples of processing techniques that are typically used to increase melt flow rate, which [the reference] cautions should *not* be used due to the risk of obtaining a broader molecular weight distribution"); *TriVascular*, 812 F.3d at 1067–69 (finding insufficient motivation to combine two references by "substitut[ing] the recessed barbs of [one reference] with the protuberances of [another reference], since [the petitioner's] proposed substitution would destroy the basic objective of the barbs"); *In re Gordon*, 733 F.2d 900, 902 (Fed. Cir. 1984) (finding no reason to modify a prior art device where the modification would render the device "inoperable for its intended purpose"); *see also Henny Penny*, 938 F.3d at 1332 (affirming the Board's finding of no motivation to combine where the Board "credited [the patent owner's] expert's testimony that following [one reference's] method of diverting and cooling the oil in [another reference's] system would introduce 'additional plumbing and complexity'"). Notably, this is not merely an economic concern over how expensive it would be to incorporate Dockser's FPP into MacMillan's PEs; the evidence shows that doing so would increase costs because of the added complexity and size of

75

IPR2021-00155
Patent 10,416,961 B2

the FPP—a concern that MacMillan explicitly counsels against. *See* Sur-Reply 14–15.

Petitioner's position is that the individual benefits of MacMillan's approach (increasing performance speed from many low-cost processors operating in parallel) and Dockser's approach (lowering power consumption using additional circuitry added to a single, more complex processor) would have motivated a person of ordinary skill in the art to combine the references' teachings by incorporating Dockser's FPP into the multiple PEs in MacMillan's device. *See* Pet. 6, 49–55. Given the express teachings of MacMillan, however, we are not persuaded that an ordinarily skilled artisan would have been motivated to combine the two different approaches in that manner. Specifically, MacMillan's repeated statements that its approach is designed to avoid increased costs from added complexity and size (and thereby allow for many low-cost processors in the device) would have dissuaded a person of ordinary skill in the art from doing so.

To the extent Petitioner and Mr. Goodin fault Patent Owner for not providing evidence supporting its position, we are not persuaded. *See* Reply 9–10; Ex. 1071 ¶¶ 30–31. Petitioner bears the burden of proving unpatentability by a preponderance of the evidence, which includes proving that a person of ordinary skill in the art would have had reason to combine the prior art references. *See* 35 U.S.C. § 316(e); *Magnum Oil*, 829 F.3d at 1380. Moreover, Patent Owner's arguments are supported by evidence. Patent Owner relies on the disclosures of Dockser and MacMillan themselves, expert testimony from Dr. Khatri (which we credit on the issue of reasons to combine Dockser and MacMillan), and cross-examination testimony from Mr. Goodin. *See* PO Resp. 35–36; Sur-Reply 14–15; Khatri Decl. ¶¶ 101–104.

IPR2021-00155
Patent 10,416,961 B2

For the reasons explained above, we are not persuaded that a person of ordinary skill in the art would have been motivated to combine the teachings of Dockser and MacMillan in the manner asserted by Petitioner. Accordingly, Petitioner has not shown, by a preponderance of the evidence, that claim 3 would have been obvious based on Dockser and MacMillan under 35 U.S.C. § 103(a).[8]

G.   Obviousness Ground Based on Dockser, Tong, and MacMillan (Claims 3 and 23)

Petitioner contends that claims 1–5, 10, 13, 14, 21, and 23–25 are unpatentable over Dockser, Tong, and MacMillan under 35 U.S.C. § 103(a). Pet. 56–60. Of those claims, claims 1, 10, and 21 are independent. Petitioner has shown that claims 1, 2, 4, 5, 10, 13, 14, 21, 24, and 25 are unpatentable. *See supra* Section II.D. Given our disposition of the grounds based on Dockser and the combination of Dockser and Tong, we do not reach Petitioner's alternative ground asserting that claims 1, 2, 4, 5, 10, 13, 14, 21, 24, and 25 are unpatentable over Dockser and MacMillan. *Boston Sci. Scimed,* 809 F. App'x at 990. Petitioner challenges claim 3 under this ground and the ground based on Dockser and MacMillan. *See supra* Section II.F. Claim 23 is challenged under this ground only. Claims 3 and 23 recite the same limitation, but depend from claims 1 and 21, respectively.

We are not persuaded that a person of ordinary skill in the art would have been motivated to combine the teachings of Dockser and MacMillan in the manner asserted by Petitioner. Thus, Petitioner has not shown, by a

---

[8] Given this determination, we need not address Patent Owner's argument that the references fail to teach the "exceeds" limitation of claim 3. *See* PO Resp. 28–35; Sur-Reply 10–14.

IPR2021-00155
Patent 10,416,961 B2

preponderance of the evidence, that the subject matter recited in claims 3 and 23 would have been obvious based on Dockser and MacMillan under 35 U.S.C. § 103(a). Our reasoning follows.

As for the unpatentability of claims 3 and 23, Petitioner presents two theories. First, Petitioner argues that it would have been obvious to operate the FPPs in the Dockser-MacMillan combination at Tong's precision levels, such that "[t]he resulting Dockser/Tong/MacMillan combination (*i.e.*, MacMillan's multi-PE system using Dockser's FPP with as low as 5-fraction-bit precision as taught by Tong)" teaches the claim limitations "for the same reasons Dockser/MacMillan does," referring to Petitioner's earlier analysis of the asserted ground based on Dockser and MacMillan. Pet. 56–57. We are not persuaded by those arguments for the reasons explained above. *See supra* Section II.F.

Second, Petitioner presents another theory based on a "customized implementation" of Dockser's FPPs. Pet. 57–60 (citing Goodin Decl. ¶¶ 389–399); Reply 15–19 (citing Ex. 1071 ¶¶ 56–62). Petitioner argues that a person of ordinary skill in the art "would have been motivated to customize Dockser's FPPs in MacMillan's PEs to only operate at precision levels lower than full FP 32-bit operations, in view of Tong's teachings that 'the fine precision of the 23-bit mantissa is not essential.'" Pet. 58 (citing Ex. 1008, 279; Goodin Decl. ¶ 393). Specifically, according to Petitioner, an ordinarily skilled artisan would have been motivated to "implement Dockser's FPPs in the embedded signal processing system with smaller than 32-bit registers to not waste circuit space or incur unnecessary cost in having some register elements that will always be unpowered" and likewise "implement the multiplier logic in Dockser's FPP to have only as many logic elements as needed to multiply mantissas of the reduced bitwidth

78

IPR2021-00155
Patent 10,416,961 B2

(smaller than 23-bit) corresponding to the precision level selected for the embedded application." *Id.* at 58–59 (citing Goodin Decl. ¶ 394).

Petitioner's analysis is premised on an "alternative interpretation" of the portion of the "exceeds" limitation referring to execution units "adapted to execute at least the operation of multiplication on floating point numbers that are at least 32 bits wide." *Id.* at 57–60. Under that interpretation, an execution unit could meet the "adapted to" clause if it is capable of 32-bit multiplication in *some* configurations. *Id.* at 59–60. Dockser's FPP, modified as described above, would not be such a unit because it has registers and multiplier logic with less than 32 bits, and thus is not capable of 32-bit multiplication in the proposed combination—rather, only the host CPU floating-point unit would meet the "adapted to" clause. *Id.* at 60.

Patent Owner responds that a person of ordinary skill in the art "would not have been motivated to re-engineer the FPP of Dockser to remove its 32-bit capability [based on Tong] and then incorporate the re-engineered Dockser unit into the system of MacMillan." PO Resp. 40–41 (citing Khatri Decl. ¶¶ 108–109). Patent Owner argues that none of the three references support Petitioner's view. *Id.* at 40–43. First, Dockser teaches a device that "support[s] a range of selectable precisions including full precision," and "[a]djusting Dockser by removing its full-precision capability violates a central tenet of Dockser—to always be able to execute full-precision operations." *Id.* at 42 (citing Khatri Decl. ¶ 114; Ex. 1007 ¶ 3). Second, Tong similarly teaches a device with "both full- and reduced-precision capabilities," and would have reinforced Dockser's teaching that "while reduced-precision might be a viable option in certain circumstances, an execution unit should retain the ability to operate at full-precision because many applications require full precision." *Id.* (emphasis

79

IPR2021-00155
Patent 10,416,961 B2

omitted) (citing Khatri Decl. ¶ 112; Ex. 1007 ¶ 3). Third, the asserted

combination, "which requires special, customized registers, logic elements,

[and] arithmetic units, . . . would increase manufacturing costs and goes

directly against the teachings of MacMillan," in particular its teaching that

adding "significant complexity" to the SIMD architecture should be avoided.

Id. at 43–44 (citing Ex. 1009, col. 24, ll. 24–26, ll. 34–36). The parties in

their Reply and Sur-Reply disagree as to the meaning of certain cited

disclosures in the references. *See* Reply 15–19; Sur-Reply 15–18.

   After reviewing the parties' arguments and evidence, and the full

disclosures of Dockser, Tong, and MacMillan, we agree with Patent Owner.

Petitioner's asserted combination is predicated on *removing* what is a

fundamental feature of Dockser and Tong—namely, the ability to operate at

full precision or select a precision less than full precision. Dockser teaches a

single processor with a default full precision mode that is capable of also

operating in selectable reduced precision modes. Ex. 1007 ¶¶ 3–4 ("[T]he

common situation is that for certain applications, . . . a reduced precision

may be acceptable, and for other applications, . . . a greater precision may be

needed. Accordingly, there is a need in the art for a floating-point processor

in which the reduced precision, or subprecision, of the floating-point format

is selectable."). Dockser's FPP operates on 32-bit full precision values and

performs arithmetic operations at a particular subprecision if the associated

"subprecision select bits" are set, but if not, operates at full precision. *Id.*

¶¶ 17–19, 23–29, 32, claims 1 (reciting "a floating-point processor having a

maximum precision" and "selecting a subprecision less than the maximum

precision"), 8, 15, 20. Dockser thus requires full precision with the

capability to select a subprecision less than full precision, as Mr. Goodin

agrees. *See* Ex. 2043, 36:16–37:17, 38:3–8 ("I don't believe Dockser

IPR2021-00155
Patent 10,416,961 B2

discloses any embodiments without floating-point . . . selectable precision."), 55:21–57:12 ("Dockser requires selectable precision"). Indeed, Petitioner explicitly relies on the fact that "Dockser's precision is selectable" when asserting that a person of ordinary skill in the art would have combined the teachings of Dockser and Tong. *See* Pet. 45; Reply 8; Sur-Reply 17. We see no basis to conclude that a person of ordinary skill in the art would have been motivated to remove the feature that Dockser describes as always present and that is the basis for Petitioner's other assertions regarding Dockser and Tong. *See Chemours*, 4 F.4th at 1377 (finding insufficient motivation to modify a prior art reference "when doing so would necessarily involve altering the inventive concept of" the reference); Public Tr. 41:21–42:8.

Tong likewise does not support Petitioner's asserted motivation to combine. Tong discloses that "not all programs need the precision provided by generic FP hardware," and "the fine precision of the 23-bit mantissa is not essential" and "a single custom FP format may be a viable option" for particular applications, but full precision is necessary for other applications. *See* Ex. 1008, 278–280. Tong states that "scientific programs such as large-scale computational fluid dynamics or electric circuit simulation . . . require a huge amount of precision." *Id.* at 279. Tong further discloses that

> by reducing either precision or range from 32 to fewer bits, we should be able to create custom FP hardware which has lower power simply because of the bit reductions. For a narrow, application-specific task, a single custom FP format may be a viable option. However, *to be more generally useful, we need to consider arithmetic architectures which can scale to different FP formats*. Even though we may be able to assume that most of our operands can be computed successfully in limited precision, it appears *inevitable that some fraction of our operands will require full IEEE-standard precision*.

81

IPR2021-00155
Patent 10,416,961 B2

Id. at 280 (emphasis added). Given those disclosures, we are not persuaded that a person of ordinary skill in the art would have been motivated to remove from Dockser's FPP the capability (i.e., selectability from full precision) that Tong characterizes as important.[9]

Finally, with respect to MacMillan, the reference states that to achieve its cost objectives, "SIMD capabilities should not add significant complexity to the architecture of a computer system for personal use." Ex. 1009, col. 5, ll. 42–44. The Dockser-Tong-MacMillan combined device would add complexity because it involves a "customized implementation" of the multiplier logic in Dockser's FPP. *See* Pet. 59–60; Reply 15. Further, to the extent some selectability remains (among levels below full precision) in Petitioner's "customized implementation," the modified Dockser FPP would still require Dockser's additional control circuitry to implement that selectable subprecision. As explained above, such circuitry would increase costs due to added complexity and size, which MacMillan expressly counsels against. *See supra* Section II.F.

---

[9] Petitioner has shown sufficiently that a person of ordinary skill in the art would have been motivated to *select* a particular precision level in Dockser's FPP based on the teachings of Tong, but not to *remove* functionality from the FPP itself. *See supra* Section II.E.2; *compare* Pet. 45 (arguing that Tong "would have motivated a [person of ordinary skill in the art] to *configure* Dockser's FPP . . . to *operate at* a *selected* precision level retaining as few as 5 mantissa fraction bits . . . or others empirically determined (using Tong's techniques) to not require greater precision" (emphasis added)), *with id.* at 58–59 (arguing that a person of ordinary skill in the art "would have been motivated to *customize* Dockser's FPPs in MacMillan's PEs" by removing register storage elements to have "smaller than 32-bit registers" (for the register bit-dropping technique) or removing multiplier logic elements (for the multiplier logic technique) (emphasis added)).

82

IPR2021-00155
Patent 10,416,961 B2

We credit the testimony of Dr. Khatri on these points, as it is explained sufficiently and is consistent with the disclosures of the references. *See* Khatri Decl. ¶¶ 46–48, 107–115. Petitioner has not shown sufficiently that a person of ordinary skill in the art would have combined the teachings of Dockser, Tong, and MacMillan to remove from Dockser's FPP the ability to select a precision less than full precision and then incorporate the modified FPP into the PEs of MacMillan. Accordingly, Petitioner has not shown, by a preponderance of the evidence, that claims 3 and 23 would have been obvious based on Dockser, Tong, and MacMillan under 35 U.S.C. § 103(a).

H. *Motion to Exclude*

Patent Owner moves to exclude (1) Exhibits 1073, 1075–1078, 1082–1084, 1088–1090, and 1096–1097 as "irrelevant and prejudicial" under Federal Rules of Evidence 402 and 403; (2) Exhibit 1082 for the same reasons and also as unauthenticated under Federal Rule of Evidence 901; (3) and Exhibits 1094 and 1095 as containing inadmissible hearsay under Federal Rule of Evidence 802. Mot. 3–7. Patent Owner's Motion is dismissed as moot, as we do not rely on the exhibits in a manner adverse to Patent Owner in our Decision.

In particular, Petitioner cites Exhibits 1073 and 1075 in support of its argument that Dockser and MacMillan teach the "exceeds" limitation of claim 8 and similar limitations in other claims. *See* Mot. 3–4; Reply 14. Petitioner cites Exhibits 1076–1078 and 1097 in response to an argument made by Patent Owner in its Response regarding MacMillan. *See* Mot. 4; Reply 16–17, 25. We need not reach those arguments because we determine that Petitioner has not shown unpatentability of the claims challenged based

IPR2021-00155
Patent 10,416,961 B2

on the combinations of Dockser and MacMillan, and Dockser, Tong, and

MacMillan, for other reasons. *See supra* Sections II.F and II.G

Petitioner cites Exhibits 1078, 1082–1084, 1088–1090, and

1094–1096 in support of its arguments regarding objective indicia of

nonobviousness. *See* Mot. 4–7; Reply 19–30. As explained above, we do

not find Patent Owner's evidence of objective indicia of nonobviousness

persuasive, and do not rely on the disputed exhibits in reaching that

conclusion.

I.   *Motions to Seal*

The parties move to seal portions of their demonstrative exhibits that

quote or describe material that we previously ordered sealed. Papers 52, 54;

see Papers 27, 48 (granting motions to seal). The parties filed unredacted

confidential and redacted public versions of their demonstrative exhibits.

Exs. 1098, 2054. Upon reviewing the materials sought to be sealed, it

appears that the parties' characterization of the materials is accurate and the

redactions are narrowly tailored to only confidential information. The

parties have established good cause to seal the redacted portions of the

demonstrative exhibits. We grant the motions to seal.

III. CONCLUSION[10]

Petitioner has demonstrated, by a preponderance of the evidence, that

claims 1, 2, 4, 5, 10, 13, 14, 21, 24, and 25 of the '961 patent are

---

[10] Should Patent Owner wish to pursue amendment of the challenged claims
in a reissue or reexamination proceeding subsequent to the issuance of this
Decision, we draw Patent Owner's attention to the April 2019 *Notice
Regarding Options for Amendments by Patent Owner Through Reissue or*

IPR2021-00155
Patent 10,416,961 B2

unpatentable, but has not demonstrated, by a preponderance of the evidence,

that claims 3 and 23 are unpatentable.

Reexamination During a Pending AIA Trial Proceeding. See 84 Fed. Reg.
16,654 (Apr. 22, 2019). If Patent Owner chooses to file a reissue application
or a request for reexamination of the challenged patent, we remind Patent
Owner of its continuing obligation to notify the Board of any such related
matters in updated mandatory notices. *See* 37 C.F.R. §§ 42.8(a)(3),
42.8(b)(2).

IPR2021-00155
Patent 10,416,961 B2

In summary:

Claims	35 U.S.C. §	References/ Basis	Claims Shown Unpatentable	Claims Not Shown Unpatentable
1, 2, 4, 5, 10, 13, 14	103(a)	Dockser	1, 2, 4, 5, 10, 13, 14	
1, 2, 4, 5, 10, 13, 14, 21, 24, 25	103(a)	Dockser, Tong	1, 2, 4, 5, 10, 13, 14, 21, 24, 25	
1–5, 10, 13, 14	103(a)	Dockser, MacMillan[11]		3
1–5, 10, 13, 14, 21, 23–25	103(a)	Dockser, Tong, MacMillan[12]		3, 23
Overall Outcome			1, 2, 4, 5, 10, 13, 14, 21, 24, 25	3, 23

---

[11] As explained above, given our disposition of the grounds based on Dockser and the combination of Dockser and Tong, we do not reach Petitioner's alternative ground asserting that claims 1, 2, 4, 5, 10, 13, and 14 are unpatentable over Dockser and MacMillan. *See supra* Section II.E.

[12] As explained above, given our disposition of the grounds based on Dockser and the combination of Dockser and Tong, we do not reach Petitioner's alternative ground asserting that claims 1, 2, 4, 5, 10, 13, 14, 21, 24, and 25 are unpatentable over Dockser, Tong, and MacMillan. *See supra* Section II.G.

IPR2021-00155
Patent 10,416,961 B2

IV. ORDER

In consideration of the foregoing, it is hereby:

ORDERED that claims 1, 2, 4, 5, 10, 13, 14, 21, 24, and 25 of the '961 patent have been shown to be unpatentable, and claims 3 and 23 of the '961 patent have not been shown to be unpatentable; and

FURTHER ORDERED that the parties' Motions to Seal (Papers 52 and 54) are *granted*, and the unredacted confidential versions of the demonstrative exhibits (Exhibits 1098 and 2054) shall remain under seal pursuant to the default protective order previously entered in the instant proceeding; and

FURTHER ORDER that Patent Owner's Motion to Exclude (Paper 46) is dismissed.

This is a final decision. Parties to the proceeding seeking judicial review of the decision must comply with the notice and service requirements of 37 C.F.R. § 90.2.

IPR2021-00155
Patent 10,416,961 B2

FOR PETITIONER:

Elisabeth H. Hunt
Richard F. Giunta
Anant K. Saraswat
Adam R. Wichman
Nathan R. Speed
WOLF, GREENFIELD & SACKS, P.C.
ehunt-ptab@wolfgreenfield.com
rgiunta-ptab@wolfgreenfield.com
asaraswat-ptab@wolfgreenfield.com

FOR PATENT OWNER:

Peter Lambrianakos
Vincent J. Rubino, III
Enrique W. Iturralde
Richard Cowell
FABRICANT LLP
plambrianakos@fabricantllp.com
vrubino@fabricantllp.com
eiturralde@fabricantllp.com
rcowell@fabricantllp.com

Brian M. Seeve
Matthew D. Vella
PRINCE LOBEL TYE LLP
bseeve@pricelobel.com
mvella@pricelobel.com